AD-A038 348 TEXAS CHRISTIAN UNIV FORT WORTH INST OF BEHAVIORAL R--ETC F/G 12/1
APPLICATIONS OF TWO-STAGE LEAST SQUARES IN CAUSAL ANALYSIS AND --ETC(U)
MAR 77 L R JAMES, B K SINGH N00014-76-C-0008

UNCLASSIFIED IBR-77-6 NL

1 OF 1
AD
A038348

END
DATE
FILMED
5-77

ADA038348

12 B.S.

D D C
APR 15 1977
C

Institute of Behavioral Research

Texas Christian University
Fort Worth, Texas 76129

(12)

Applications of Two-Stage Least Squares in
Causal Analysis and Structural Equations

Lawrence R. James & B. Krishna Singh

Completed under Office of Naval Research
Contract Number NONR N00014-76-C-0008
S. B. Sells, Ph.D., Principal Investigator

IBR Report No. 77-6 March, 1977

Applications of Two-Stage Least Squares in

Causal Analysis and Structural Equations

Lawrence R. James and B. Krishna Singh

Institute of Behavioral Research

Texas Christian University

Running head: Two-Stage Least Squares

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

14

1. REPORT NUMBER: IBR-77-6

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

9

4. TITLE (and Subtitle): 6 Applications of Two-Stage Least Squares in Causal Analysis and Structural Equations,

5. TYPE OF REPORT & PERIOD COVERED: Technical Report

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s): 10 James, Lawrence R. Singh, B. Krishna

8. CONTRACT OR GRANT NUMBER(s): 15 ~~NONR~~ N00014-76-C-0008 New

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Institute of Behavioral Research
Texas Christian University
Fort Worth, TX 76129

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

12 86 p.

11. CONTROLLING OFFICE NAME AND ADDRESS
Organizational Effectiveness Research Programs
Office of Naval Research (Code 452)
Arlington, VA 22217

12. REPORT DATE: 11 Marc[redacted]77

13. NUMBER OF PAGES: 83

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report): Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; Distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Causal Analysis
Cross-Lagged Panel Correlations
Instrumental Variables
Reduced Forms
Structural Equations
Structural Parameters
Two-Stage Least Squares

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A procedure, two-stage least squares (2SLS), for analyzing structural equations when one or more explanatory variables are correlated with the error or disturbance term is reviewed. A brief introduction to structural equations and the use of ordinary least squares in causal analysis is presented initially. This is followed by an introduction to 2SLS and the application of 2SLS to designs in which (a) two or more variables are reciprocal causes of each other, (b) one or more variables contain random

next page

180 820

cont

measurement error, and (c) lagged values of one or more dependent variables are used as predictors. A major objective for the review of these particular applications of 2SLS is to demonstrate how salient problems in psychology can be addressed by use of structural equations.

Abstract

A procedure, two-stage least squares (2SLS), for analyzing structural equations when one or more explanatory variables are correlated with the error or disturbance term is reviewed. A brief introduction to structural equations and the use of ordinary least squares in causal analysis is presented initially. This is followed by an introduction to 2SLS and the application of 2SLS to designs in which (a) two or more variables are reciprocal causes of each other, (b) one or more variables contain random measurement error, and (c) lagged values of one or more dependent variables are used as predictors. A major objective for the review of these particular applications of 2SLS is to demonstrate how salient problems in psychology can be addressed by use of structural equations.

Applications of Two-Stage Least Squares in Causal

Analysis and Structural Equations

A procedure, two-stage least squares (2SLS), for estimating structural parameters when one or more explanatory variables are correlated with the error or "disturbance" term is reviewed. A brief introduction to structural equations and the use of ordinary least squares in causal analysis is presented initially. This is followed by an introduction to 2SLS and the application of 2SLS to designs in which (a) two or more variables are reciprocal causes of each other, (b) one or more variables contain random measurement error, and (c) lagged values of one or more dependent variables are used as predictors, which focuses on the analysis of the cross-lagged panel correlation design in terms of structural equations. The above applications were selected because they were presumed to be of interest to psychologists. They are not, however, exhaustive of the applications of 2SLS.

STRUCTURAL EQUATIONS AND CAUSAL ANALYSIS

A structural equation refers to the "representation of the true structural or causal properties of real-world phenomena, as contrasted with equations that are merely used for prediction or estimation purposes" (Namboodiri, Carter, & Blalock, 1975, p. 448). Some recent efforts in psychology have attempted to acquaint psychologists with causal inferences based on recursive structural equations (cf. Feldman, 1975; Kalleberg & Kluegel, 1975; Kenny, 1975; Kerlinger & Pedhazur, 1973; Sims & Szilagyi, 1975; Werts & Linn, 1970). A recursive design is one in which the hypothesized causal relationships are unidirectional or asymmetric, as demonstrated in Figure 1. A recursive model requires that in a causal sequence each X_i must precede each X_j, where

$\underline{i} < \underline{j}$, and that X_j may be caused either directly or indirectly by each X_i, but that an X_i cannot be caused by an X_j. Thus, causal closure implies that the relationships are asymmetric in a recursive or unidirectional causal model.

Insert Figure 1 about here

Recursive models are applicable to numerous types of data sets (cf. Werts & Linn, 1970), but there exist many designs for which they are not. Examples where recursive models would not suffice are provided by social system theory (cf. Indik, 1968; James & Jones, 1976; Katz & Kahn, 1966; Lichtman & Hunt, 1971; Sells, 1963, 1968) and interactional paradigms (cf. Bowers, 1973; Ekehammar, 1974; Endler, 1975; Endler & Magnusson, 1976). These theories project complex models that incorporate feedback and reciprocal causation or simultaneity (Goldberger, 1973; Singh & Williams, 1972; Singh, 1975), and include reciprocal interactions between individuals and situations over time (Overton & Reese, 1973). Another example is a postulate of system theory which states that all events are correlated, thus obscuring unidirectional cause-effect relationships (Katz & Kahn, 1966). These theories raise obvious questions for structural analyses that employ recursive models and their accompanying structural equations. Rather, nonrecursive models that incorporate reciprocal causation are required. An example of a nonrecursive model is presented in Figure 2. In this model, an X_j may be caused directly or indirectly by several X_i, including cases where $\underline{i} < \underline{j}$.

Insert Figure 2 about here

It will be emphasized throughout this paper that the structural equations representing recursive or nonrecursive structural models, or variations of these models such as block-recursive models (cf. Fisher, 1966), can be formulated in terms of the general linear model. When viewed in this manner, structural equations may be employed to test causal hypotheses for a multitude of designs, including (a) truly experimental designs involving randomization and intervention (cf. Miller, 1971), (b) quasi-experimental designs, and (c) cross-sectional (static) correlational designs involving data obtained from natural settings. Applications of structural equations to different designs are of course based upon different assumptions, some of which may not be testable. Furthermore, the assumptions are of primary importance because simply casting analyses in terms of structural equations does not guarantee that one is actually testing causal hypotheses.

Our attention here is focused on applications of structural equations to correlational designs which employ "passive data" (Cook & Campbell, 1976) obtained on either static or longitudinal bases. It is important to emphasize that the application of structural equations to correlational data, particularily from cross-sectional designs, does not represent an attempt to identify an unambiguous, unique causal model. Rather, it represents a method for examining the logical consistency of alternative causal hypotheses and models, and for rejecting those that are untenable (Duncan, 1966; Feldman, 1975; Goldberger, 1973; Kerlinger & Pedhazur, 1973; Werts & Linn,

1970). When there are a priori, untestable assumptions or different structural equations that fit the data equally well, a single, correct set of structural equations cannot be ascertained. However, untenable causal models that do not fit the data may be discarded. Thus, the application of structural equations to correlational data does not represent an attempt "to accomplish the impossible task of deducing causal relations from the values of correlation coefficients" (Wright, 1934, in Duncan, 1966, p. 15), but an attempt to examine whether any alternative theoretical models accepted for possible causal interpretation are logically consistent with the data.

This should not be construed to mean that one is necessarily "making causal inferences based on correlational data", which is both ambiguous and an overstatement (Duncan, 1975, p. 47). On the other hand, the unfortunate result of rejecting the consideration of causal relations in correlational designs has frequently led to descriptive rather than explanatory interpretations of data, or even worse, to degradation of research designs to the extent that subgoals such as maximizing validity coefficients in the absence of explanatory theory have become ends in themselves. By contrast, thinking in terms of structural equations and causality with nonexperimental data requires a strong theoretical base and has definite advantages. Among these are an emphasis on explanation rather than description (Strotz & Wold, 1971), estimation of change rather than fixed values (Namboodiri et al., 1975), and a pattern of interpretation that makes explicit the rationale and assumptions underlying analytical procedures while simultaneously forcing the discussion of results to be at least internally consistent (Duncan, 1966). Thus, explanatory theories encompassing change and causality, and the goodness

of fit of data to such theories, become the primary focus of structural equations. If nothing else, such an approach forces investigators to place primary concern on theoretical issues.

To state the matter directly, it is possible to adopt causal, structural equation models with correlational data based on natural observations if one is willing to make assumptions regarding prior time sequencing of variables and relationships among variables (some of which may not be testable) and statistical assumptions (some of which may also not be testable). The fact that economists, and more recently sociologists, have been routinely employing such models is a case in point. The adoption of such models and accompanying assumptions provides the basis for the remainder of this paper.

STRUCTURAL EQUATIONS AND THE USE OF ORDINARY LEAST SQUARES

Structural equations are forms of the general linear model, and when appropriate, may be represented as ordinary least squares (OLS) multiple regression equations, where standardized or unstandardized (partial) regression weights provide unbiased and consistent estimates, based on a sample, of the population causal or structural parameters. That is, when regression weights are employed as structural, rather than just statistical, parameters, then reference is being made to causal, real-world phenomena (Duncan, 1975; Heise, 1975; Namboodiri et al., 1975). For example, an unstandardized regression weight employed as an estimate of a structural parameter provides an indication of "the mean change in the dependent variable expected to result for each unit of change in one particular independent variable, assuming other independent variables are held constant" (Darlington & Rom, 1972, p. 452). On the other hand, it is possible to demonstrate several

sources of error which result in inconsistent and biased parameter estimation if OLS is employed. The use of OLS for parameter estimation is discussed here briefly. This is followed by a discussion of conditions which preclude the use of OLS, and necessitate the use of other procedures, particularly 2SLS.

In general, in a causal system one is usually addressing several dependent variables. If a structural equation for each dependent variable is delineated, a system of simultaneous equations is obtained. Generally, the system of simultaneous equations may be viewed as a system of multiple regression equations, where the direct causal factors for each dependent variable are considered predictors. If one (structural) equation is selected from this system, it would take the general form (assuming linearity and additivity)

$$Y_g = a + b_{g1} X_1 + \ldots + b_{gk} X_k + \ldots + b_{gK} X_K + d_g \qquad (1)$$

where Y_g represents the dependent variable, a represents the intercept, $X_1 \ldots X_K$ represent the predictors (raw scores), $b_{g1} \ldots b_{gK}$ represent the unstandardized regression coefficients for the X_k, and d_g represents the disturbance term.[1] (Disturbance terms are synonymous with error terms, and include variance resulting from sampling error as well as effects of unknown or unmeasured outside influences such as measurement error and variables that assist in causal explanation but are omitted from the theory and/or measurement).

In this context, the b_{gk} coefficients represent sample estimates of population structural parameters. It will be assumed that the predictors

are random variables and that the samples are random and large. Given these conditions, one of the primary concerns is whether the $\underline{b_{gk}}$ provide consistent estimates of the population structural parameters. It will be assumed that if random samples of increasing size ($\underline{n}$) are selected, then estimates of the structural parameters ($\underline{b_{gk}}$) will converge to the population structural parameters ($\underline{B_{gk}}$). This is referred to as a probability limit, or "plim", and is designated by: $\text{plim}\ \underline{b}^{(\underline{n})} = \underline{B}$ (Johnston, 1972), which connotes that $\underline{b}$ will converge to $\underline{B}$ in the limit. Consistent estimators may or may not be asymptotically unbiased, although in most applications consistent estimators also tend to be asymptotically unbiased estimators (Pindyck & Rubinfeld, 1976, p. 24).

The assumptions required to apply OLS to equation 1 to obtain estimates of the structural parameters include linearity, additivity, interval scales, and random sampling. It is also generally assumed that the disturbance terms are distributed as $\underline{N}\ (0, \sigma_{\underline{d}}{}^2)$. Additional assumptions required for at least consistent estimation are:

(1) the predictors have no measurement error (random or nonrandom),

(2) the disturbance terms from different structural equations are uncorrelated, and

(3) the predictors are uncorrelated in the limit with the disturbance term. This assumption may be represented as $\text{plim}\ [(1/\underline{n})\ \underline{X}\ \underline{d}] = 0$ (cf. Christ, 1966; Johnston, 1972; Theil, 1971), and implies that any variable which is causally connected to the dependent variable but has been left out of the structural equation is not causally connected to any of the predictors (cf. Blalock, Wells, & Carter,

1970). That is, the effects of an omitted variable will be part of the disturbance term, therefore resulting in a correlation between the disturbance term and any predictor causally related with the omitted variable. From another perspective, this assumption implies that all major causes of the dependent variable are included in the structural equation.

Any violation of the above assumptions may lead to biased and inconsistent estimates of the structural parameters if OLS is employed. Of primary importance is the last assumption, which is almost impossible to meet in empirical research. That is, unless the theoretical causal system is completely known and all predictors measured with perfect reliability, the last assumption will ordinarily be violated and some degree of inconsistency and bias will be introduced into the structural parameter estimates. Problems associated with omitted variables will be addressed in more detail later; however, at this time other conditions resulting in correlations between predictors and disturbance terms, and the necessity for using procedures other than OLS for consistent parameter estimation, will be discussed.

The first of these conditions involves a situation where one or more of the predictors is in fact a dependent variable in another equation in the system (e.g., $\underline{X}_1 = \underline{Y}_{g+1}$), and the two dependent variables ($\underline{Y}_g$ and $\underline{Y}_{g+1}$) are reciprocally related. At least part of the equation system is then nonrecursive, the variable that is reciprocally related to the criterion is correlated with the disturbance term, and therefore OLS is not appropriate for parameter estimation. The second condition arises when one or more of

the predictors include random measurement errors. This condition also results in a correlation between the predictors measured with error and the disturbance term, and again OLS is not appropriate for parameter estimation. Finally, if one of the predictors is in fact a lagged value of the dependent variable (e.g., $X_2 = Y_{gt-1}$), and the disturbance terms are serially correlated over time, then the lagged dependent variable will be correlated with the disturbance term in the limit and OLS should again not be employed.

Each of the above three conditions is addressed separately in this report. It will be shown how 2SLS, or modified 2SLS, can be applied to the structural equations to obtain at least consistent estimates of the structural parameters (assuming that other assumptions are met). These applications are discussed in the following order: (a) analysis of nonrecursive structural equations and an introduction to 2SLS, (b) the application of 2SLS to structural equations which include predictors that have random measurement error, and (c) the application of a modified version of 2SLS to structural equations which include lagged dependent variables and serially correlated disturbances.

THE ANALYSIS OF NONRECURSIVE STRUCTURAL EQUATIONS AND AN INTRODUCTION TO 2SLS

Logic of Nonrecursive Models

A nonrecursive model is one in which two or more variables to be explained by the model (i.e., dependent variables) are mutually dependent and reciprocal causes of one another. It is also assumed that the mutual effects of the two or more variables in mutual interaction are relatively rapid or at least the time lags are short and cannot be meaningfully identified nor

measured (cf. Namboodiri et al., 1975) (If meaningful time lags are identifiable and measureable, then the model can be treated as recursive). For exemplary purposes, we shall assume that the models presented here incorporate only cross-sectional, correlational data based on natural observations, and that the structural equations are algebraic in form. It is also assumed that the relationships among the mutually interacting variables are stable, or have reached an equilibrium - type condition (Miller, 1971); this is discussed in more detail below.

The nonrecursive model selected for this discussion is presented in Figure 3. In this figure, the three variables labeled by a Y (i.e., Y_1, Y_2, Y_3) are endogenous variables. Endogenous variables are dependent measures that are to be explained by the theory or model. For example, as shown in Figure 3 each of the endogenous variables is dependent upon each of the other endogenous variables through a system of reciprocal relationships. Each endogenous variable is also dependent upon one or more predetermined variables, which are represented by the variables labeled by an X (i.e., X_1, X_2, X_3, X_4). In general, predetermined variables consist of (a) lagged values of the endogenous variables, and (b) exogenous variables, which are lagged or non-lagged variables that are considered to be separate causes of the endogenous variables. The predetermined variables are treated as "givens" and are assumed to provide explanatory power to the model but are not themselves to be explained by the model. Moreover, they are not dependent on the endogenous variables, and are treated as predictors or independent variables. Because Figure 3 is a cross-sectional model, the predetermined variables (X_k) consist only of non-lagged exogenous variables.

Insert Figure 3 about here

The curved lines among the exogenous variables in Figure 3 connote that relationships exist among these measures. Curved lines also indicate relationships that are not explained by the model. As shown in Figure 3, X_1, X_2, and X_3 are intercorrelated, but X_4 is not correlated with any of the other exogenous variables. The arrows in the model, both from exogenous variables to endogenous variables and among the endogenous variables (i.e., reciprocal relationships), represent the causal inferences. Associated with each arrow is an unstandardized regression coefficient or structural parameter (e.g., b_{12}). Finally, each endogenous variable has associated with it a disturbance term, which is designated by small "d" and subscripted by a numeral corresponding to the numeral of the endogenous variable.

Returning briefly to the assumption of equilibrium, it is assumed that for each subject the mutual effects of the three endogenous variables (Y_1, Y_2, Y_3) on each other have reached a state of stability, and the levels of each of the variables are constant for each individual subject. It is further assumed that the levels of the exogenous variables are temporally fixed for each subject, that the effects of the exogenous variables on the endogenous variables have been relatively rapid, and that the structural model is appropriate for all members of the sample (and population). It is then possible to conduct the analysis by employing comparisons across subjects to infer processes that have been at work <u>within</u> subjects (Namboodiri et al., 1975).[2]

Statistics of Nonrecursive Models -- An Introduction to 2SLS

A number of statistical procedures are available for the analysis of nonrecursive models, including indirect least squares, 2SLS, three-stage least squares, and limited-information and full-information maximum likelihood functions. We shall focus here on 2SLS, which has been shown to be applicable to social science data (cf. Duncan, 1970; Duncan, Haller, & Portes, 1968, 1971; Kohn & Schooler, 1973; Mason & Halter, 1968; Miller, 1971; Namboodiri et al., 1975; Waite & Stolzenberg, 1976), and is generally considered to be as powerful as more sophisticated methods (cf. Christ, 1966; Johnston, 1972; Theil, 1971; King, Note 1). The introduction to 2SLS begins with a discussion of identification, proceeds to statistical assumptions, analytical procedures, and tests for goodness of fit, and concludes with a summary in which an application of a nonrecursive model and 2SLS is proposed for a current psychological research problem.

Two-stage least squares, developed separately by Baseman (1957) and Theil (1953a,b), is a simultaneous equation estimation method in which the estimation of structural parameters is conducted independently for each equation in the system. However, 2SLS cannot proceed without first addressing the question of <u>identification</u>, which involves determination of whether sufficient information exists to estimate the unknown structural parameters of the structural equations (Theil, 1971). It should be noted that structural parameters and structural equations are population terms; in the presentation below, however, we continue to use the term structural equations for data based on random samples, but differentiate between population structural parameters and their sample estimates.

Identification

Theil (1971, pp. 448, 449) has succintly defined identification in the following manner

> . . . when we have a complete linear system of *L* equations, the parameters of the *j*th equation are not estimable when there exists a linear combination of the other *L* - 1 equations that does not contain any of the variables of the system which do not occur in the *j*th equation; or, to put it in more positive terms, the parameters of the *j*th equation are not estimable when there exists a linear combination of the other equations that contains only the variables which do occur in the *j*th equation, and possibly fewer. In that case the *j*th equation is said to be not identifiable (or underidentified) in its system.

In the general case, and using present terminology, the identification of an equation rests on meeting two conditions, which are the order condition and the rank condition. With respect to the order condition, there will be a system of *G* structural equations representing *G* endogenous variables. To these equations will be added *K* predetermined variables, which as noted earlier will consist only of non-lagged exogenous variables in the present examples. For identification purposes, each equation must have *G* - 1 variables deleted from the total possible *G* + *K* variable set, where the deleted variables may be either endogenous or exogenous. If the number of variables deleted from an equation is equal to *G* - 1, the equation is exactly identified. If the number of variables deleted is greater than *G* - 1, the equation is overidentified. Otherwise, the equation is underidentified, and no solution

exists. It should be noted that within a set of G equations, some equations may be exactly identified while others may be overidentified (or even under-identified).

While the order condition is necessary, it is not sufficient for the identification of an equation. A necessary and sufficient condition is the rank condition (cf. Fisher, 1966). A discussion of the rank condition involves considerable mathematical complexity, and thus such discussion was included in a technical appendix to this paper (Appendix B). However, it is generally the case that if the order condition is met, the rank condition is also met (Namboodiri et al., 1975). An exception occurs when the structural equations for two or more endogenous variables contain the same combinations of variables.

It is important to note that the selection and addition of exogenous (or more generally, predetermined) variables to the structural equations should be based on sound theory, and that trivial variables should not be added to the equations solely for identification purposes (Duncan, 1975). The criteria for selection of exogenous variables are that a) the hypothesized direct effects should be significant and substantial, b) the hypothesized indirect effects should be significant and substantial, and c) the exogenous variables are not dependent on the endogenous variables at the time of the study.

Statistics of 2SLS

The use of 2SLS is illustrated by an example, using the nonrecursive model presented in Figure 3.

Design of the structural equations. The first step in 2SLS is to write out the structural equations for each of the endogenous variables. The structural equation for *each endogenous variable includes* those endogenous and exogenous variables and their (estimated) parameters that have a direct relationship (i.e., arrows in Figure 3) with the endogenous variable, plus the disturbance term. The structural equations for Figure 3, written in deviation form, are

$$y_1 = b_{12}\, y_2 + b_{13}\, y_3 + c_{11}\, x_1 + c_{12}\, x_2 + d_1 \qquad (2)$$

$$y_2 = b_{21}\, y_1 + b_{23}\, y_3 + c_{23}\, x_3 + d_2 \qquad (3)$$

$$y_3 = b_{31}\, y_1 + b_{32}\, y_2 + c_{34}\, x_4 + d_3 \qquad (4)$$

where the b_{gh} ($g \neq h$) represent estimates of the structural parameters for the mutually interacting endogenous variables, the c_{gk} represent estimates of the structural parameters for the exogenous variables, and the d_g represent disturbance terms.

Order condition. With respect to the order condition required (but not sufficient) for identification, $G = 3$, $K = 4$, and $G + K = 7$. The first equation, equation 2, is exactly identified because $G - 1 = 2$ variables have been *deleted from the equation* (i.e., there are five variables, including the dependent variable, in the equation). Equations 3 and 4 are overidentified because more than two variables have been deleted from each equation.

Statistical assumptions. The statistical assumptions underlying the 2SLS procedure are:

1) The causal effects are linear and additive.

2) Variables have been measured on interval scales.

3) The independent variables have no random nor nonrandom measurement error, which in the above model would mean that all variables should be perfectly reliable because each endogenous variable is used as an independent variable (as well as a dependent variable).

4) The exogenous variables ($\underline{X_k}$) are uncorrelated with the disturbance terms ($d_{\underline{g}}$) in the limit (i.e., plim $[(1/\underline{n})\, X_{\underline{gk}}\, d_{\underline{g}} = 0]$). As noted earlier, this implies that all major causes of the dependent variables have been ascertained.

5) $\underline{E}\,(d_{\underline{g}}) = 0$, and the disturbances are normally distributed (an assumption that allows the use of statistical tests [cf. Johnston, 1972]).

6) The sample selected is random if drawn from a finite population.

An additional assumption is that the variables are ordered correctly. This is similar to the identification question (i.e., some variables are deleted from each equation) and implies that selection of variables is based on theory and hopefully involves previous research (Duncan, 1975). Violations of the above assumptions are generally referred to as specification errors (cf. Spaeth, 1975).

Of importance here is the omission of the assumptions associated with recursive models and OLS that all variables in an equation be uncorrelated in the limit with the disturbance term of that equation, and that the disturbance terms of different equations be uncorrelated. In nonrecursive models, the endogenous, but not the exogenous, variables are assumed to be correlated with disturbance terms, and the disturbance terms may also be correlated. The rationale for these conditions is that the mutual interactions among a set of endogenous variables result in influences on each endogenous variable

by the disturbance terms of the other endogenous variables (cf. Johnston, 1972, p. 343). That is, the disturbance terms include variance representing reciprocal causes of the endogenous variables which are reciprocally related (Namboodiri et al., 1975).

With respect to the assumptions of 2SLS, the absence of measurement error or the restriction of the equations to variables measured by interval scales might appear to be overly confining to psychologists. Fortunately, models are available for guiding analysis when some of the assumptions cannot be met. Models developed for random measurement error are discussed in the following section of this paper, and the reader is referred to Namboodiri et al. (1975) for a discussion of models which involve nonrandom measurement error. With respect to interval scales, while some authors have argued that ordinal scales will suffice for parametric purposes (cf. Bohrnstedt & Carter, 1971; Spaeth, 1975), it appears reasonable to require that the scales be "essentially interval" (i.e., while perhaps not perfectly interval, the scales should possess interval qualities and be regarded as substantially better than ordinal). This rationale is based on the causal interpretation of a structural parameter presented earlier, where one would presume that a unit increase in the predictor should connote (approximately) uniform degrees of change throughout the range of the scale (the same is true for dependent variables). On the other hand, paradigms are available for the use of ordinal and nominal scales in structural equations (cf. Boyle, 1970; Namboodiri et al., 1975). Such paradigms are closely associated with present statistical knowledge in psychology inasmuch as structural equations are a form of the general linear model.

Violations of assumptions regarding linearity and additivity have also been addressed (cf. Darlington & Rom, 1972; Namboodiri et al., 1975; Werts & Linn, 1970). Because structural equations are forms of the general linear model, polynomial regression and the use of cross-products to represent interaction terms are applicable, although there are some questions as to whether these procedures can be employed with random variables (Sockloff, 1976). Moreover, the addition to the structural equations of squared, cubed, etc., terms and possible moderators and cross-products involving moderators (cf. Saunders, 1956) may result in problems concerning identification, multicollinearity, and interpretation of the regression weights for cross-product terms. Identification problems pertain to the need to include more predetermined variables in the equations when polynomial regression or interaction terms increase the number of endogenous variables. Multicollinearity concerns the problem where highly intercorrelated variables (e.g., a variable and a cross-product term in which the variable is included) lead to "bouncing beta weights" and large sampling errors for the estimates of the structural parameters (Darlington, 1968; Werts & Linn, 1970). The multicollinearity problem is not limited to polynomial regression and interaction analysis; it can occur with any highly correlated variables that enter into the same equation. Methods for alleviating multicollinearity include deletion of variables, formation of composites, and factor analysis (cf. Goldberger, 1971; Johnston, 1972; Jöreskog, 1970). Finally, the interpretation of regression weights for cross-product terms is questionable because "regression weights in nonlinear regression equations can be changed by changing the means of the independent variables, and the means are often chosen arbitrarily" (Darlington

& Rom, 1972, p. 453). The reader is referred to the Darlington and Rom paper for possible solutions to this problem.

In practice, the assumption concerning the lack of correlation between the exogenous variables and the disturbance terms in the limit is often violated, as noted earlier with OLS. The omission of relevant variables from the model and spuriousness are of particular concern. It is generally impossible to assume that all relevant variables are known and included in an equation for a particular endogenous variable (cf. Duncan, 1975; Heise, 1975; Kenny, 1975; Spaeth, 1975). The costs associated with omitting relevant variables are a function of their importance in the system and of the way in which their effects are transmitted throughout the system (Spaeth, 1975). The effects of omitted variables might include a) biased estimates of at least some of the structural parameters included in the model and underestimation of the dependent, endogenous variable (Duncan, 1975; Spaeth, 1975); b) alternative explanations of results based on spurious relationships between measured and unmeasured common causes (Kenny, 1975) and, as noted above, c) correlations among predetermined variables and the disturbance terms, as well as correlations among the disturbance terms for reasons other than simultaneity (cf. Miller, 1971).

In general, the omitted variable specification error can be quite serious because it implies an incomplete theoretical system. On the other hand, presently unknown omitted variables might be responsible for the specification error, or it may be difficult to obtain reliable and accurate measures of variables of hypothesized theoretical importance (e.g., an ultimate criterion). In practice, therefore, it is not uncommon to allow

certain trade-offs. For example, exogenous variables with low rather than zero correlations with the disturbance term in the limit may be accepted (cf. Fisher, 1971). We note here, however, that in the statistical presentation below the exogenous variables are assumed to be uncorrelated with the disturbance terms in the limit.

Analytical procedures. The presumed correlations between the endogenous variables and the disturbance terms in the limit (i.e., plim $[(1/\underline{n})\ Y_{\underline{g}}\ \underline{d}_{\underline{h}}] \neq 0$, where $\underline{g} \neq \underline{h}$) result in inconsistencies and bias if OLS is used to estimate the values of the structural parameters in nonrecursive models. The 2SLS procedure is employed to obtain estimates of the endogenous variables that are uncorrelated with the disturbance terms in the limit for the equations in which the endogenous variables are used as predictors. These estimates are then used to obtain consistent estimates of the structural parameters. Thus, two stages of estimation are required to estimate the structural parameters. A discussion of the general algebraic steps involved in applying 2SLS to nonrecursive equations is presented below. An outline of the matrix algebra steps involved in this procedure and a discussion of the rank condition required for identification are presented in Appendix B.

To obtain estimates of the endogenous variables that are uncorrelated with the disturbance terms, a reduced form of the set of structural equations is constructed. A reduced form, which is the first-stage of 2SLS, consists of a set of equations in which each endogenous variable is represented as a function of only the exogenous (predetermined) variables and a disturbance term (Duncan, 1975). That is, each endogenous variable serves as dependent variable for one equation, and the independent variables for each of these

equations include all exogenous (predetermined) variables from the system of equations, plus a disturbance term. Ordinary least squares is then applied to each of the reduced form equations to obtain estimates of each of the endogenous variables (see Appendix A for an example derivation of a reduced form).

For example, the reduced forms for the model in Figure 3, and equations 2 through 4, are

$$\hat{y}_1 = \hat{\pi}_{11} x_1 + \hat{\pi}_{12} x_2 + \hat{\pi}_{13} x_3 + \hat{\pi}_{14} x_4 + m_1 \quad (5)$$

$$\hat{y}_2 = \hat{\pi}_{21} x_1 + \hat{\pi}_{22} x_2 + \hat{\pi}_{23} x_3 + \hat{\pi}_{24} x_4 + m_2 \quad (6)$$

$$\hat{y}_3 = \hat{\pi}_{31} x_1 + \hat{\pi}_{32} x_2 + \hat{\pi}_{33} x_3 + \hat{\pi}_{34} x_4 + m_3 \quad (7)$$

In equations 5 through 7, the variables are presumed to be in deviation form, the $\hat{y}_g$ represent predicted scores for the endogenous variables, the $\hat{\pi}_{gk}$ represent unbiased estimates of population reduced form parameters (π_{gk}) based on OLS in a random sample, and the m_g represent disturbance terms for the reduced form equations.

It is important to note that the predicted y_g ($\hat{y}_g$) are exact functions of the x_k (exogenous variables) and thus the correlations between the $\hat{y}_g$ and the disturbance term in a particular structural equation (i.e., equations 2 through 4) are equal to zero in the limit (cf. Johnston, 1972, p. 383). That is, the x_k are uncorrelated with the disturbance terms in the limit and therefore exact functions of the x_k will also be uncorrelated with the disturbance terms in the limit. In a sample the $\hat{\pi}_{gk}$ are estimates of the π_{gk} and may not result in predicted y_g that have zero sample correlations with

the respective disturbance terms, although divergences from zero tend to be smaller as sample sizes increase (Duncan, 1975).

Assuming that the correlations between the $\hat{y}_g$ and d_g are at least asymptotically equal to zero, it is possible to proceed to the second-stage of the 2SLS procedure, which involves replacing the original sample values of the y_g with the $\hat{y}_g$ in equations 2 through 4 and conducting OLS. The new equations for estimating the structural parameters, based on a random sample, are

$$y_1 = \tilde{b}_{12}\,\hat{y}_2 + \tilde{b}_{13}\,\hat{y}_3 + c_{11}\,x_1 + c_{12}\,x_2 + d_1' \tag{8}$$

$$y_2 = \tilde{b}_{21}\,\hat{y}_1 + \tilde{b}_{23}\,\hat{y}_3 + c_{23}\,x_3 + d_2' \tag{9}$$

$$y_3 = \tilde{b}_{31}\,\hat{y}_1 + \tilde{b}_{32}\,\hat{y}_2 + c_{34}\,x_4 + d_3' \tag{10}$$

where the c_{gk} for the exogenous variables are unchanged with respect to equations 2 through 4, but the $\tilde{b}_{gh}$ for the endogenous variables indicate that these estimates of the structural parameters are based on predicted y_g rather than original y_g.

The regression weights provided by equations 8 through 10 are consistent estimators of the population structural parameters (B_g), but they are not generally unbiased, although the bias tends to become negligible in large samples (cf. Johnston, 1972; Namboodiri et al., 1975). The parameter estimates for a given set of structural equations are mathematically, but not necessarily causally, unique if the equations are exactly identified or overidentified. The significance of the estimated structural parameters can be tested using well-known significance tests for unstandardized (partial)

regression coefficients (cf. Kerlinger & Pedhazur, 1973). The null hypothesis is that the population structural parameter does not differ significantly from zero (i.e., has no causal effect). In overidentified models it will be noted, however, that one or more variables are presumed to have structural parameters equal to zero in particular equations. As shown below, this provides an avenue for assessing the goodness of fit of the data to the model, where in fact the estimates of certain parameters may change (thus questioning the uniqueness in a causal sense of the parameter estimates in the same, overall causal model).

Goodness of fit. As discussed earlier, the goal of causal analyses based on correlational data is to examine the logical consistency of alternative causal hypotheses and to reject those that are untenable. However, because of untestable assumptions or sets of structural equations that fit the data equally well, a simple, "correct" set of structural equations cannot be ascertained. With respect to the present example, it is quite possible that different models and therefore different sets of structural equations could be developed. In fact, in the typical case a rather considerable number of alternative models can be constructed (cf. Duncan, 1975).[3] For example, the reciprocal relationship between Y_1 and Y_3 could be replaced with a single arrow from Y_1 to Y_3. Thus, the theoretical assumptions and the questions of the goodness of fit of a particular set of structural equations to a set of data become salient. A method for testing the goodness of fit is presented, which can be conducted only with overidentified equations.

The test of goodness of fit discussed here is estimation of omitted parameters. Other, and often more sophisticated procedures are available

(cf. Costner & Schoenberg, 1973; Kalleberg & Kluegel, 1975; Jöreskog, 1973; Namboodiri et al., 1975), but they generally require stronger assumptions and the differences among the methods are not definitive (Johnston, 1972; Namboodiri et al., 1975; King, Note 1).

In the example above, it will be recalled that structural equations 3 and 4 were overidentified. In essence this means that in these equations certain variables were assumed to have population structural parameters equal to zero. For example, in equation 3, the population structural parameters $\underline{C_{21}}$, $\underline{C_{22}}$, and $\underline{C_{24}}$ were assumed to be zero. A test of the goodness of fit based on sample data would be to ascertain empirically if in fact at least some of the estimates of these parameters (i.e., $\underline{c_{21}}$, $\underline{c_{22}}$, $\underline{c_{24}}$) are equal to zero. As outlined by Namboodiri et al. (1975), the parameters in which there is the least faith of a zero value are inserted into the structural equations until each structural equation is exactly identified. In the example, only one sample structural parameter could be inserted into equations 3 and 4 to achieve exact identification. Once the overidentified equations have been exactly identified, a 2SLS analysis is conducted on the new set of structural equations. Significance tests for all estimated structural parameters are then conducted. Particular interest is attached to (a) whether the estimated structural parameters which the causal model specified as being equal to zero are in fact "approximately" zero (i.e., within the realm of sampling error), and (b) any meaningful changes in the original estimates of the structural parameters when compared to the first 2SLS analysis with overidentified models.

The correlations between the exogenous (predetermined) variables and the disturbances terms, as well as the correlations among the disturbance terms from different equations, may also be checked (see Duncan, 1975 for computing equations). Although the latter set of correlations is not constrained to equal zero, large correlations would bring the model into question. For example, a large correlation between two disturbance terms could indicate the presence of omitted variables that should be in the model, or of correlated, nonrandom errors.

<u>Summary</u>

Theoretical and mathematical developments in this section are explained, by way of summary, first by postulating how a nonrecursive model might be applied to a salient problem in social-organizational psychology, and then by reviewing verbally the steps involved in employing 2SLS for analysis purposes. The problem selected concerns the causal determinants of leader and subordinate behaviors in formal groups (e.g., workgroups).

A basic assumption underlying much of the leadership research has been that the behavior of the leader toward subordinates is a major causal factor in respect to organizationally related attitudes and behaviors of subordinates (cf. Gibb, 1969; Kerr & Schriesheim, 1974; Likert, 1967; Stogdill, 1974; Vroom, 1976). A considerable body of data provide support for this assumption, which reflects an asymmetric or recursive model. However, an increasing accumulation of research has indicated that leader behavior is at least partially determined by the behaviors of subordinates and by leader-subordinate relationships, and further that a particular leader may display different (flexible) behaviors with different subordinates in the same workgroup

(cf. Barrow, 1976; Cummins, 1972; Dansereau, Graen, & Haga, 1975; Evans, 1973; Farris & Lim, 1969; Fiedler & Chemers, 1974; Green, 1973, 1975; Hill & Hughes, 1974; House & Mitchell, 1974; Lowin & Craig, 1968).

The latter type of relationship is illustrated by studies involving experimental or quasi-experimental designs (including cross-lagged panel correlation) in which high or increasing subordinate performance levels caused leaders to employ more supportive-consideration oriented behaviors, while low or decreasing subordinate performance levels resulted in more use of structuring-authoritarian behaviors on the part of the leader (Barrow, 1976; Dansereau et al., 1975; Green, 1973, 1975). Other studies, primarily of a correlational nature, have suggested that a number of subordinate variables might affect supervisory behavior, either directly or as moderators. These include (a) job knowledge, (b) satisfaction, (c) role ambiguity and role conflict, (d) locus of control (e.g., internals were more satisfied with considerate leader behaviors), (e) race of both supervisor and subordinate, (f) perceived organizational independence, (g) hierarchical level of subordinate in the organization, (h) expectations concerning leader behavior and rewards, (i) the acceptance of the leader by subordinates, (j) needs for structure and independence (and the degree of congruency between leaders and subordinates for these needs), (k) complexity of subordinate tasks, and (l) various other needs such as needs for achievement and performing meaningful tasks (cf. Herold, 1974; House & Mitchell, 1974; Kerr & Schriesheim, 1974; Lowin & Craig, 1968; Parker, 1976; Steers, 1975; Stogdill, 1974; Vroom, 1976). There are, of course, a number of additional contingencies, such as the leader's hierarchical influence, specificity of goals, power and authority,

organizational incentives and feedback, organizational structure, and others that might well enter into the determination of leader-subordinate relationships.

This is not an exhaustive review and it is realized that the leadership process involves many contingencies. Nevertheless, there is ample evidence to postulate that the causal relationships between leader and subordinate behaviors are at least partially symmetric and reciprocal rather than asymmetric. Furthermore, it appears that the effects of many of the reciprocal interactions between leaders and subordinates are relatively rapid and thus that meaningful time lags are essentially unidentifiable. Finally, if we assume that the reciprocal relationships between leader and subordinate behaviors tend to stabilize in situations, then a nonrecursive model is appropriate for attempting to identify the causal factors for both leader and subordinate behaviors.

The following steps provide a rough outline of the application of 2SLS for the analysis of the proposed nonrecursive relationships between leader and subordinate behaviors. For exemplary purposes, we shall assume that (a) the data are cross-sectional and based on natural observations; (b) the unit of analysis is the subordinate, where the managerial strategies (Oldham, 1976) employed by each leader for each subordinate are measured, while other data describing the leader are duplicated for each subordinate; (c) moderator analyses would be conducted using the subgrouping technique (cf. Guion, 1976) (this avoids the potential problem of multicollinearity if cross-product terms were used, although it would require the construction of separate and perhaps different structural equations for each subgroup); (d) the structural

equations are identified; and (e) the structural equations are otherwise correctly specified (e.g., assumptions regarding linearity, reliability, random sampling, inclusion of all relevant causal factors, and so forth have not been violated). The structural parameters for each dependent variable for which a reciprocal relationship exists could then be estimated by the following steps (these steps follow roughly those presented by Heise [1975, p. 169]).

1. Design a structural equation for each dependent variable that expresses the values of the dependent variable as a function of other endogenous variables with which the dependent variable has reciprocal relationships, predetermined variables (nonlagged exogenous in this case) with which there is a direct relationship, and a disturbance term. For our purposes here, we shall assume that there are two <u>sets</u> of endogenous variables, namely (a) leader managerial strategies (e.g., providing rewards and punishments, setting goals, designing feedback systems, etc.), and (b) subordinate behaviors (e.g., job performance levels on different behavioral criteria, reactions to the leader, etc.). Reciprocal relationships are presumed to exist between the sets of endogenous variables for at least one variable from each set, and, if appropriate, for selected variables within each set.[4]

It is also assumed that there are three sets of exogenous variables, namely (a) variables that have direct effects on leader behavior and indirect effects (through leader behavior) on subordinate behavior (e.g., leader intelligence, experience, hierarchical influence, power, authority, etc.); (b) variables that have direct effects on subordinate behavior and indirect effects on leader behavior (e.g., subordinate intelligence, experience,

knowledge, satisfaction, motivation, etc.); and (c) variables that have direct effects on both leader and subordinate behaviors (e.g., structure of the organization, subsystem, and workgroup, complexity of the tasks, congruency indices between leader and subordinate needs, expectations, and race, etc.). Thus, as an example, the structural equation for a particular subordinate behavior (dependent variable) would include the following predictors (a) leader behaviors which have a reciprocal relationship with the dependent variable, (b) other subordinate behaviors which have a reciprocal relationship with the dependent variable, and (c) exogenous variables that have direct effects on the dependent variable, which would include variables that directly affect only subordinate behaviors as well as variables that directly affect subordinate behaviors and leader behaviors.

2. Separate from the system of all equations those variables which are exogenous. These variables cannot include a variable with which the reciprocally related endogenous variables (i.e., leader and subordinate behaviors) have a reciprocal relationship.

3. Regress, using OLS, each of the reciprocally related endogenous variables on all of the exogenous variables identified in step 2 to obtain regression equations for predicting values of the endogenous variables (i.e., develop a reduced form and conduct the first-stage regression). Use the first-stage regression equations to obtain predicted values for the reciprocally related endogenous variables. These predicted values will be purged, in the limit, of their correlations with the disturbance terms associated with the structural equations constructed in step 1.

4. Return to the structural equations constructed in step 1 and estimate the structural parameters by OLS, substituting the predicted values of reciprocally related endogenous variables obtained in step 3 for the original values of the endogenous variables (i.e., conduct the second-stage regression).

At this point, the 2SLS procedure and hypotheses regarding parameter estimates and goodness of fit can be addressed following procedures discussed earlier.

RANDOM MEASUREMENT ERROR

Random measurement errors, particularily in the predetermined variables, may have disturbing effects on the estimation of structural parameters. A general treatment of the effects of random measurement error on parameter estimation, which is often referred to in econometrics as the "error in variables" problem, is described below. This is followed by an introduction to the use of instrumental variables as a solution to the random measurement error problem, and a demonstration of the relationships between instrumental variables and 2SLS. It will also be noted that 2SLS is the more general procedure because it can be used with overidentified equations.

From a general standpoint, a bivariate relationship between two variables in a random sample may be displayed in deviation form as

$$\underline{y} = \underline{b}\,\underline{x} + \underline{d} \qquad (11)$$

where $\underline{b}$ is an estimate of population structural parameter $\underline{B}$, $\underline{x}$ is a random variable which takes on values from a distribution of true scores randomly sampled from the population, and $\underline{d}$ represents the disturbance term.

For purposes of unbiased and consistent estimation using OLS, it is assumed that the explanatory variable ($\underline{x}$) is uncorrelated with the disturbance term in the limit [i.e., plim [(1/$\underline{n}$) $\underline{X}$ $\underline{d}$ = 0]. However, it can be shown rather easily that this assumption is incorrect if $\underline{x}$ involves a random error component (only random measurement error is addressed here). That is, if the observed $\underline{x}$ is equal to $\underline{t}$ + $\underline{e}$, where $\underline{t}$ equals the true score on the variable and $\underline{e}$ is a random measurement error, then equation 11 becomes

$$\underline{y} = \underline{b}\,\underline{x} + (\underline{d} - \underline{b}\,\underline{e}) \qquad (12)$$

which was obtained by replacing $\underline{x}$ in equation 11 with $\underline{x}$ - $\underline{e}$ (The $\underline{x}$ in equation 11 assumes no measurement error, or conversely, a true score. Thus, if the observed $\underline{x}$ is measured with error, the term in equation 11 should be $\underline{x}$ - $\underline{e}$ [=t]).

In equation 12, $\underline{x}$ is correlated with the disturbance term ($\underline{d}$ - $\underline{b}$ $\underline{e}$) in the limit because $\underline{x}$ is a function of $\underline{e}$ (cf. Blalock et al., 1970; Bohrnstedt, 1969; Christ, 1966; Goldberger, 1971; Johnston, 1972; Theil, 1971; Wiley & Wiley, 1971). Thus, the use of OLS to estimate $\underline{B}$ from $\underline{b}$ will be both biased and inconsistent (random measurement error in $\underline{y}$ is considered a part of the $\underline{d}$ term and does not affect bias or consistency of parameter estimation). In the bivariate case, the bias is in the direction of attenuating (underestimating) the estimate of $\underline{B}$. However, in the multivariate case with several explanatory variables, each with different random measurement errors, the bias in the estimates of the structural parameters may be postive or negative. For example, as discussed by Blalock et al. (1970), it is possible (a) to infer that a relationship between two variables is partly spurious when in fact it is totally spurious, (b) to treat an additive model as if a

statistical interaction existed, and (c) to obtain incorrect estimates of structural parameters, including a sign reversal (cf. Kenny, 1975), particularily when the explanatory variables are correlated and have differing degrees of random measurement error.

A number of authors have proposed approaches for dealing with random measurement error in variables, especially if the observed measures are considered as fallible "indicators" of unobserved constructs (e.g., true scores) (cf. Blalock, 1969, 1970; Blalock et al., 1970; Bohrnstedt, 1969; Costner, 1969; Duncan, 1975; Goldberger, 1971; Goldberger & Duncan, 1973; Hauser & Goldberger, 1971; Heise, 1975; Johnston, 1972; Jöreskog, 1973; Kalleberg & Kluegel, 1975; Kenny, 1975; Namboodiri et al., 1975; Pindyck & Rubinfeld, 1976; Werts & Linn, 1970; Werts, Linn, & Jöreskog, 1971; Wiley, 1973; Wiley & Wiley, 1971; Wold, 1975). We shall focus here only on observables, and employ a popular approach known in econometrics as "instrumental variables". The relationship between instrumental variables and 2SLS will be demonstrated. It should be noted, however, that the "multiple indicator" procedures, which are not addressed here, have a strong tie to known methods in psychology (e.g., confirmatory factor analysis, multitrait-multimethod matrix) and thus offer another attractive alternative to the analysis of variables with random measurement error and to the analysis of unobservables.

The discussion of the instrumental variables approach and its relationship to 2SLS generally follows presentations by Heise (1975), Johnston (1972), and Pindyck and Rubinfeld (1976). Beginning with equation 12, where $\underline{x}$ is correlated with the disturbance term, the instrumental variables approach proceeds by attempting to find a variable "$\underline{z}$" such that (cf. Heise, 1975)

(1) plim $[(1/n)\ Z\ (d - b\ e)] = 0$; which connotes that z is not causally connected to factors which affect y but have been omitted from the equation, nor is z related to the random measurement error in x.

(2) z is a cause for x, preferably a direct cause although an indirect cause through intervening variables is acceptable as long as the intervening variables are not causally related to y. In addition, z itself cannot affect y directly. (If z is causally related to y in any way other than through x, then it is possible for z to be causally related to omitted causes for y and thus create a specification error, such as violating the preceeding assumption).

(3) z is not affected causally by y or x.

(4) z may have random measurement error as long as such error is not correlated with the disturbance term in equation 12 (a highly reliable z is of course preferable).

Given these conditions, with accompanying assumptions of linearity, additivity, essentially interval measurement, random sampling, and $E\ (d - b\ e) = 0$, it is possible to use z to obtain a consistent estimate of B by the following equation

$$\frac{\Sigma\ y\ z}{\Sigma\ x\ z} \qquad \text{(sum is from 1 . . . . n)} \qquad (13)$$

which replaces the usual OLS estimation: $b = (\Sigma\ x\ y)\ /\ (\Sigma x^2)$; and where z is used as an instrument for x.

In the above equation, b will be a consistent estimator of B because

$$\text{plim } \underline{b} = \underline{B} + \frac{\Sigma\, \underline{z}\, (\underline{d} - \underline{b}\, \underline{e})}{\Sigma\, \underline{x}\, \underline{z}}$$

where the numerator for the second term on the right-side of the equation approaches 0 in the limit (see assumption 1).

The rationale for equation 13 can be seen more clearly if the intuitive and statistical relationships between instrumental variables and 2SLS is demonstrated. In essence, equation 13 is represented by the model $\underline{z} \rightarrow \underline{x} \rightarrow \underline{y}$ (Heise, 1975), where $\underline{x}$ is measured with error. Assuming that $\underline{z}$ meets the criteria for an instrumental variable, the first-stage of 2SLS consists of replacing the fallible measure ($\underline{x}$) with an estimate ($\hat{\underline{x}}$) that is not correlated with the disturbance term ($\underline{d}$ - $\underline{b}$ $\underline{e}$) in the limit. In this context, the first-stage of 2SLS involves the creation of an instrument (Pindyck & Rubinfeld, 1976). The second-stage of 2SLS then involves the use of the created instrument ($\hat{\underline{x}}$) in place of the fallible measure to obtain a consistent estimate of the structural parameter. For example, the first-stage of 2SLS consists of regressing $\underline{x}$ on $\underline{z}$ and obtaining an estimated $\hat{\underline{x}}$. The regression equation is

$$\hat{\underline{x}} = \underline{a}\, \underline{z} + \underline{m}$$

where the predicted score (instrument) for $\underline{x}$ ($\hat{\underline{x}}$) is equal to $\underline{a}$ $\underline{z}$, which by definition is not correlated with ($\underline{d}$ - $\underline{b}$ $\underline{e}$) in the limit. The second-stage of 2SLS then consists of replacing $\underline{x}$ with $\hat{\underline{x}}$ in equation 12, and conducting OLS, although a more direct comparison of 2SLS and instrumental variables is shown by the following simple derivation

For the usual OLS calculation of $\underline{b}$, (Σ $\underline{y}$ $\underline{x}$) / ($\Sigma \underline{x}^2$), the $\underline{x}$ in the numerator is replaced by $\hat{\underline{x}}$, and one $\underline{x}$ in the denominator is replaced

by $\hat{\underline{x}}$ (replacing both $\underline{x}$'s in the denominator with $\hat{\underline{x}}$ results in an inconsistency [cf. Christ, 1966; Pindyck & Rubinfeld, 1976]). We thus have

$$\underline{b} = \frac{\Sigma \underline{y}\ \underline{x}}{\Sigma \underline{x}\ \underline{x}} = \frac{\Sigma \underline{y}\ (\underline{a}\ \underline{z})}{\Sigma \underline{x}\ (\underline{a}\ \underline{z})} = \frac{\Sigma \underline{y}\ \underline{z}}{\Sigma \underline{x}\ \underline{z}}$$

where the last term is the same as the instrumental variable estimator presented in equation 13.

In this example, 2SLS is equivalent to instrumental variables. This will not always be the case; the instrumental variables approach typically focuses on the use of only one instrument for each fallible variable. Where more than one instrument exists for a fallible variable, each instrument provides a separate estimate for the structure parameter (e.g., equation 13 is replicated for each instrument), and the problem is then to decide which estimate to accept, or how to combine the separate estimates to arrive at one estimate (cf. Goldberger, 1971, 1973). On the other hand, 2SLS automatically accommodates multiple instruments for each fallible variable because a least squares weighted combination of multiple instruments ($\underline{z}_p$) can be employed to create a new instrument ($\hat{\underline{x}}$) for each fallible variable in the first-stage regression. In other words, the first-stage of 2SLS provides the basis for developing a weighted linear combination of the original instruments in the creation of a new instrument.

In more general terms, 2SLS and the instrumental variables approach will provide unique and identical parameter estimates in an exactly identified system of simultaneous equations if all predetermined variables are used in the first-stage of 2SLS and the instruments (analogous to $\hat{x}$) used in

the second-stage of 2SLS consist of the predicted values from the first-stage (reduced form) regressions (Pindyck & Rubinfeld, 1976). In overidentified structural equations, however, 2SLS again provides unique parameter estimates while the instrumental variables approach does not. Furthermore, as noted by Goldberger (1973, p. 151), 2SLS "is as efficient as any other instrumental variable estimator in the present context", which referred to the weighting and combination of instruments in overidentified structural equations.

The application of 2SLS to simultaneous equations where some variables involve random measurement errors is summarized below, and follows the presentation by Johnston (1972). To simplify matters, it was assumed that the equations were based on a recursive model, although nonrecursive or block-recursive models could also be treated in this general paradigm (i.e., both nonrecursiveness and random measurement error would have to be addressed as reasons for correlations among explanatory variables and disturbance terms). Matrix algebra was employed to conserve space (the reader may wish to consult Appendix B before proceeding with this section).

To begin, one equation from the system of simultaneous equations is represented by

$$\underset{\sim}{y} = \underset{\sim}{Y_1}\,\underset{\sim}{b} + \underset{\sim}{X_1}\,\underset{\sim}{c} + \underset{\sim}{u} \tag{14}$$

where $\underset{\sim}{y}$ in an $\underline{n}$ x 1 vector of observations (raw scores) for an endogenous variable,

$\underset{\sim}{Y_1}$ is an $\underline{n}$ x $\underline{g}$ matrix of observations on variables which include random measurement errors and are correlated with the disturbance term ($\underset{\sim}{Y_1}$ does not include $\underset{\sim}{y}$),

b is a $g \times 1$ vector of structural parameter estimators attached to the Y_1 variables,

X_1 is an $n \times k$ matrix of observations on variables appearing in this equation which are not correlated with the disturbance term,

c is a $k \times 1$ vector of structural parameter estimators attached to the X_1 variables, and

u is an $n \times 1$ vector of disturbances, which can also be written as $(d - E\,b)$, where d is the vector of original disturbance terms and $E\,b$, analogous to $b\,e$ equation 12, represents the effects of random measurement errors in the Y_1 variables.

If it is presumed that the equations are identified and all assumptions for 2SLS have been met, with the exception of the random measurement errors in the Y_1 variables, then the specification error for the above equation is

$$\text{plim}\ (\tfrac{1}{n}\, Y_1'\, u) \neq 0$$

which connotes that the variables in Y_1 are correlated with the disturbance term in the limit, and further that the estimates b of the structural parameters B will be biased and inconsistent.

It is important to reiterate that the variables in X_1 are not correlated with u in the limit. In addition, there will exist a set of variables in the remainder of the simultaneous equations that are also not correlated with the disturbance terms. These variables do not appear in X_1, and will be considered as comprising a matrix X_2. (Although we are not dealing with a nonrecursive model here, perhaps an analogy to such a model would be of assistence. That is, the Y_1 variables are analogous to the interdependent endogenous variables in nonrecursive models in the sense that they are correlated with the

disturbance terms. The X_1 and X_2 variables are analogous to predetermined variables in the sense that they are not correlated with the disturbance terms).

The application of 2SLS to the above situation is designed to develop a set of instruments for the Y_1 variables in the first-stage regression which are purged of their correlations with the disturbance term, and then to use these instruments in place of Y_1 in the second-stage regression in order to obtain consistent estimates of the structural parameters. The reduced form employed in the first-stage regression for the estimation of the Y_1 variables is based on both the X_1 and X_2 variables, which by definition are not correlated with the disturbance term. Thus, estimates of Y_1 (i.e., $\hat{Y}_1$), which are direct linear functions of X_1 and X_2, will also not be correlated with the disturbance term. The reduced form, first-stage regression is therefore (Johnston, 1972, p. 381):

$$\hat{Y}_1 = X\,(X'X)^{-1}\,X'Y_1 \tag{15}$$

where $X = [X_1 \quad X_2]$.

The second-stage regression proceeds by first noting that $\hat{Y}_1 = Y_1 - V_1$, where V_1 is an $n \times g$ matrix of the residuals obtained from regressing Y_1 on X. Second, in matrix terminology the instruments used in the second-stage regression are comprised by the matrix $[Y_1 - V_1 \quad X_1]$, which will be employed in place of $[Y_1 \quad X_1]$, the original observation matrix (the reader will note that X_1 does not change in the above two matrices and that $Y_1 - V_1$ provides a computing method for $\hat{Y}_1$ which precludes the need to actually determine the values for $\hat{Y}_1$ [Johnston, 1972, pp. 382 and 390]. Given these conditions, the estimating equations for the second-stage of 2SLS are:

$$\begin{bmatrix} (\underset{\sim}{Y}_1 - \underset{\sim}{V}_1)' \underset{\sim}{Y}_1 & (\underset{\sim}{Y}_1 - \underset{\sim}{V}_1)' \underset{\sim}{X}_1 \\ \underset{\sim}{X}_1' \underset{\sim}{Y}_1 & \underset{\sim}{X}_1' \underset{\sim}{X}_1 \end{bmatrix} \begin{bmatrix} \tilde{\underset{\sim}{b}} \\ \underset{\sim}{c} \end{bmatrix} = \begin{bmatrix} (\underset{\sim}{Y}_1 - \underset{\sim}{V}_1)' \underset{\sim}{y} \\ \underset{\sim}{X}_1' \underset{\sim}{y} \end{bmatrix} \quad (16)$$

The above computing equations will provide consistent estimates of the structural parameters as long as

$\text{plim}\,[\frac{1}{n} (\underset{\sim}{Y}_1 - \underset{\sim}{V}_1)' \underset{\sim}{u}'']$ and $\text{plim}\,(\frac{1}{n} \underset{\sim}{X}_1' \underset{\sim}{u}) = 0$ (cf. Johnston, 1972; Pindyck & Rubinfeld, 1976).

In summary, an outline of the use of instrumental variables and 2SLS in situations where some explanatory variables are correlated with the disturbance term for reasons of random measurement error has been presented. Space limitations preclude a thorough discussion of specification errors which are salient for the random measurement error problem; however, we shall briefly mention some of the more important of these errors (it should also be noted that all assumptions regarding the use of 2SLS are operational).

Of initial concern are the criteria mentioned earlier for the selection of instrumental variables. For example, as discussed by Fisher (1971) and Blalock et al. (1970), it is often difficult to obtain instrumental variables which are uncorrelated in the limit with the disturbance term while at the same time being major and direct causes for the fallible variables. Moreover, the use of multiple instruments for one variable (i.e., in overidentified equations) may result in the problem of multicollinearity. In the former case, it may be necessary to use instrumental variables which have low rather than zero correlations with the disturbance term in the limit, and/or are indirect causes for the explanatory variables comprised partially

of random measurement errors. In the latter case, some instrumental variables may have to be deleted or a principal components analysis conducted on the instrumental variables in order to obtain independent predictors (Amemiya, 1966; Johnston, 1972).

Another set of concerns pertains to the decision of whether to use instrumental variables and/or 2SLS versus OLS when the criteria for instrumental variables have not been fully met. For example, Blalock et al. (1970) demonstrated that the use of an instrumental variables approach (and 2SLS) may be inferior to OLS if an instrumental variable is related to the dependent variable either directly or indirectly. Finally, as noted earlier the multiple indicator approaches, which focus on unobservables (e.g., confirmatory factor analysis), provide another avenue for addressing variables with random measurement errors, and perhaps a combination of the procedures discussed here and the multiple indicator approaches might well provide the most viable methods of analysis for the random measurement error problem.

In conclusion, it must be stressed that the procedures described in this section do not provide a solution when unreliable variables are used. For example, one could question the efficacy of the predicted scores following the first-stage regression if the variables to be replaced were highly unreliable to begin with. Rather, it is presumed that the variables to be replaced are at least moderately reliable. Furthermore, it is also questionable that instruments should be developed for variables that are highly, but not perfectly, reliable (cf. Blalock et al., 1970). That is, the specification errors associated with developing instrumental variables (e.g., multicollinearity) might have more serious contaminating effects on parameter

estimation than simply proceeding with highly reliable variables and recognizing that some bias and inconsistency might be present in estimation. In fact, Duncan (1975) has noted that highly reliable variables encompassing small amounts of random measurement errors would not be likely to provide undue strain on structural models.

LAGGED ENDOGENOUS VARIABLES

The present application of 2SLS addresses the question of dynamic analysis, namely the inclusion of lagged explanatory variables in structural equations. As noted earlier, both endogenous and exogenous variables may be lagged. Furthermore, the model may be recursive or nonrecursive, or a combination of both recursive and nonrecursive (e.g., block-recursive). In the application of 2SLS selected for discussion here, the rather thorny problem of including lagged endogenous variables in nonrecursive structural equations is presented. This application provided an opportunity to demonstrate how two applications of 2SLS might be combined (i.e., nonrecursive models and lagged endogenous variables); however, the application is not exhaustive of the applications of 2SLS for structural models with lagged variables. For example, Johnston (1972, pp. 318-320) has presented a procedure which includes a version of 2SLS for analyzing recursive models with lagged endogenous variables.

Another reason for selecting this application of 2SLS was its implications for a currently popular procedure employed for causal analysis in psychology, namely the cross-lagged panel correlation design (Campbell, 1963; Cook & Campbell, 1976; Campbell & Stanley, 1963; Feldman, 1975; Kenny, 1973, 1975). In fact, it is hoped that this discussion will serve to encourage psychologists

to begin to think of cross-lagged panel correlation designs in terms of their place in a more holistic theoretical system as well as in terms of competing causal hypotheses (e.g., nonrecursive rather than recursive models). With this interest in mind, the application of 2SLS for nonrecursive structural equations which include lagged values of one or more endogenous variables was addressed by formulating the analysis of the cross-lagged panel correlation design in terms of structural equations.

The reader is referred to Kenny (1975) for a review of the cross-lagged panel correlation (XLPC) design. We shall focus here on a brief comparison of the goals of XLPC and structural equations, and then proceed to the application of structural equations to the XLPC design. As shown in Figure 4, the XLPC design involves two dependent or endogenous variables measured at the _same_ time ($\underline{Y_{1t}}$ and $\underline{Y_{2t}}$, where _t_ represents observations), and two lagged values of the endogenous variables, both measured at time _t-1_ ($\underline{Y_{1t-1}}$ and $\underline{Y_{2t-1}}$). The latter variables are considered predetermined in the present context. The XLPC analysis is a test for spuriousness, the null hypotheses being that the relationship between $\underline{Y_{2t}}$ and $\underline{Y_{1t-1}}$ (for example) is due to the effects of one or more other variables rather than causal relationships between the two variables (Kenny, 1975). However, failure to reject the null hypothesis is not sufficient to conclude that the relationship was in fact spurious (cf. Kenny, 1975).

Insert Figure 4 about here

As noted by Kenny, the XLPC design is intermediary, in terms of causal explanation, between purely correlational designs and well-elaborated structural models. Kenny further reported that although XLPC and structural models have been contrasted, the two models address different objectives and make somewhat different assumptions. For example, the XLPC design is a test for spuriousness, does not require that all causal variables be included in the model, and does accommodate measurement error. Structural equations, on the other hand, focus on the estimation of causal parameters, and, as noted earlier, omitted causal variables and measurement error result in specification errors. Thus, structural equations are typically more demanding, both in terms of theory and psychometric/statistical criteria. Kenny also concluded that XLPC designs were more applicable to social science data, given the present state of theoretical systems and the pragmatics of measurement. While we agree with this conclusion, we also feel that a great deal is to be gained by thinking in terms of more complete theoretical systems, identifying sources of spuriousness (i.e., omitted variables), and improving upon measurement techniques. For these reasons, the application of structural equations to XLPC designs was addressed, with the presumption that such applications represent a desirable goal for psychology.

A number of authors have proposed methods for transforming the XLPC design into structural (or path) equations (cf. Duncan, 1969; Bohrnstedt, 1969; Goldberger, 1971; Heise, 1970; Pelz & Lew, 1970). These equations typically involve lagged endogenous variables, including (a) lagged values of the dependent variable (e.g., $\underline{Y_{1t-1}}$ is viewed as a cause for $\underline{Y_{1t}}$) and (b) what will be referred to here as "cross-lagged endogenous variables" (e.g., $\underline{Y_{1t-1}}$

is viewed as a cause for Y_{2t}). Furthermore, as discussed later it can frequently be assumed that serial correlation exists among the disturbance terms for structural equations representing an endogenous variable measured at different points in time. Such conditions require special and somewhat complex statistical procedures for solution, including several modified versions of 2SLS (cf. Amemiya, 1966; Fair, 1970; Fisher, 1971; Johnston, 1972; Miller, 1971; Nerlove, 1971; Pindyck & Rubinfeld, 1976). The procedure presented by Fair (1970) was recommended by Pindyck & Rubinfeld (1976) as an optimal solution for structural equations involving lagged endogenous variables, and was used as basis for this presentation.

In constructing the structural equations for the XLPC design and lagged endogenous variables, the possibility of a nonrecursive relationship between the dependent variables was added to the model presented in Figure 4 (i.e., a reciprocal interaction between Y_{1t} and Y_{2t}). Although XLPC designs typically rule out the possibility of nonrecursive causal relationships between the dependent variables "by fiat" (Cook & Campbell, 1976), their inclusion in the model provided a more general discussion of lagged endogenous variables while at the same time attending to the concerns of several authors that such relationships may be meaningful, competing hypotheses for XLPC (Duncan, 1969; Goldberger, 1971; Heise, 1970). The structural equations for the XLPC design were therefore

$$y_{1t} = b_{12} y_{2t} + c_{13} y_{2t-1} + c_{14} y_{1t-1} + d_{1t} \tag{17}$$

$$y_{2t} = b_{21} y_{1t} + c_{24} y_{1t-1} + c_{25} y_{2t-1} + d_{2t} \tag{18}$$

where all variables are presented in deviation form.

In each equation, the endogenous variable is seen as a function of (a) the other endogenous variable measured at both time _t_ (a nonrecursive relationship) and time _t-1_ (a cross-lagged, first-order autoregressive relationship), and (b) a lagged value of the endogenous variable (a lagged, first-order autoregressive relationship). All of the variables with c_{gk} regression coefficients on the right-side of equations 17 and 18 are regarded as predetermined. For the present purposes, the following assumptions were made: (1) linearity and additivity, (2) essentially interval measurement, (3) no measurement errors, (4) $E\ (d_{gt}) = 0$, and (5) random sampling. In addition, it was assumed that: (6) the model followed a first-order autoregressive scheme with discrete time lags (this allowed the use of difference rather than differential equations), (7) the variables were measured at the same points in time (i.e., synchronicity [cf. Kenny, 1975]), (8) the measurement intervals corresponded to the causal intervals, and (9) the structural relationships were invariant with respect to time (i.e., stationarity [cf. Kenny, 1975; Pindyck & Rubinfeld, 1976]).

As with earlier applications, a number of the above assumptions are difficult to meet in research. Problems associated with the first four assumptions have been discussed, while problems regarding the latter set of assumptions are discussed in a number of publications cited earlier which deal with XLPC designs, autocorrelation, and/or time-series analyses. We shall focus here on a third set of assumptions which intrinsically cause estimates of the structural parameters in structural equations 17 and 18 to be inconsistent. In general, these assumptions may be categorized as follows: (a) the disturbance terms for each dependent variable are likely to be

serially correlated over time (the disturbance terms d_{1t} and d_{2t} are also likely to be correlated because of the nonrecursive relationship and because of serial correlation among the disturbance terms), and (b) all predetermined variables are likely to be correlated with one or both disturbance terms. Each of these assumptions is addressed below. It should also be noted that equations 17 and 18 are underidentified.

10. The current values of the endogenous variables are correlated in the limit with the disturbance terms of the equations in which they are used as predictors.

That is:

$$\text{plim } (\tfrac{1}{n} Y_{2t} d_{1t}) \neq 0, \text{ and plim } (\tfrac{1}{n} Y_{1t} d_{2t}) \neq 0$$

This is due to the nonrecursive nature of the design.

11. The disturbance terms for the current and lagged endogenous variables are most likely serially correlated. This problem can be visualized by means of Figure 5, which employed Heise (1970) as a base. Y_{1t-2} and Y_{2t-2} represent an assumed, but not actually measured, additional wave of data. Potentially estimable relationships given two waves of data are depicted by solid lines, although because equations 17 and 18 are underidentified, no causal relationships could be estimated until additional predetermined variables are added. Dashed lines delineate implied relationships, and the curved line between Y_{1t-1} and Y_{2t-1} denotes that these variables are being treated as predetermined in equations 17 and 18 (with two waves of data, the relationship between these variables is estimable but not causal).

Insert Figure 5 about here

As an example, if a stable, omitted variable exists which is a cause for Y_{1t}, Y_{1t-1}, and Y_{1t-2}, then the disturbance terms d_{1t}, d_{1t-1}, d_{1t-2} will be serially correlated (i.e., the stable, omitted variable, which is part of the disturbance terms, correlates with itself over time). This problem can be avoided by including all relevant causal variables in the equations so that the disturbance terms reflect only random and unstable influences. Because equations 17 and 18 include only a few variables, it is likely that stable, causal variables have been omitted, thus resulting in a serial correlation among the disturbance terms.

The correlations among the disturbance terms for times t and $t-1$ may be represented as

$$d_{1t} = p_{11}\, d_{1t-1} + \varepsilon_{1t} \text{ and } d_{2t} = p_{22}\, d_{2t-1} + \varepsilon_{2t}$$

where p_{11} and p_{22} represent first-order serial correlation coefficients which vary between 1 and -1 (if p is greater than $|\pm 1|$, the system explodes), and ε_{1t} (and ε_{2t}) is a random error component, distributed N (0, σ_{ε}^{2}), and is independent of other disturbances for Y_{1t} (Y_{2t}) measured at different points in time, including d_{1t} (d_{2t}) (cf. Pindyck & Rubinfeld, 1976).

Although not discussed here, tests for serial correlation of disturbances in the presence of lagged endogenous variables are presented in the econometrics literature (cf. Johnston, 1972, pp. 312-313).

Several procedures are available for removing the serial correlation from the disturbance terms, including first-order differencing, generalized differencing, and generalized least squares estimation. The generalized differencing

and generalized least squares procedures are approximately equivalent for first-order serial correlation (Johnston, 1972); the former technique is briefly described here and the reader is referred to econometrics texts for more extensive treatments of both procedures. In general, the generalized differencing process replaces p, which is generally unknown, with an estimated value (processes for estimation are not discussed here), and then replaces each term in the structural equation with a difference score based upon the estimated p times a first-order lagged variable. For example, equation 17 would be

$$y_{1t} - \hat{p}_{11}\, y_{1t-1} = b_{12}\,(y_{2t} - \hat{p}_{11}\, y_{2t-1}) + c_{13}\,(y_{2t-1} - \hat{p}_{11}\, y_{2t-2}) + c_{14}\,(y_{1t-1} - \hat{p}_{11}\, y_{1t-2}) + (d_{1t} - \hat{p}_{11}\, d_{1t-1})$$

where $\hat{p}_{11}$ represents an estimate of p_{11} (a population term), and an additional wave of data would have to be obtained (i.e., Y_{1t-2} and Y_{2t-2}).

As discussed later, if $\hat{p}_{11}$ is a "good" estimate of p_{11}, then the disturbance term for the above equation would be ε_{1t}, the random error component. That is, $d_{1t} - \hat{p}_{11}\, d_{1t-1}$ will be equal to ε_{1t}, and serial correlation will have been removed from the disturbance term.

12. If the disturbance terms are serially correlated, then the lagged values of the endogenous variables will be correlated in the limit with the disturbance terms for both equations 17 and 18. That is

$\text{plim}\,(\frac{1}{n}\, Y_{1t-1}\, d_{1t})$, $\text{plim}\,(\frac{1}{n}\, Y_{2t-1}\, d_{2t})$, $\text{plim}\,(\frac{1}{n}\, Y_{1t-1}\, d_{2t})$, and

$\text{plim}\,(\frac{1}{n}\, Y_{2t-1}\, d_{1t})$ are all not equal to 0.

For example, unmeasured causes of Y_{1t-1} are related to unmeasured causes of Y_{1t} through the paths $Y_{1t-1} \longleftarrow d_{1t-1} \longrightarrow d_{1t}$ in Figure 5, thus building in a correlation between Y_{1t-1} and d_{1t}. Moreover, following the above logic, paths can be used to show a relationship between Y_{1t-1} and d_{2t} (or Y_{2t-1} and d_{1t}).

13. It will be assumed that following the generalized differencing process discussed above, the lagged values of the endogenous variables (Y_{1t-s}, Y_{2t-s}, $s = 1. . .S$) will not be correlated *in the limit* with either ε_{1t} or ε_{2t} (Fair, 1970). This provides a basis for consistent and asympotically efficient estimation of the structural parameters, although the estimates will be biased in small samples (Johnston, 1972). However, due to the nonrecursive nature of the model, the conditions described in assumption 10 for the correlations between the current values of the endogenous variables and the disturbance terms will still be in effect for ε_{1t} and ε_{2t} (e.g., Y_{2t} will still be correlated in the limit with ε_{1t}).

In summary, OLS estimates of the structural parameters for equations 17 and 18 will be biased and inconsistent because the model is nonrecursive, the disturbances are serially correlated, and the predetermined variables are correlated with the disturbances. In addition, as noted earlier, the structural equations are underidentified. The identification problem is addressed first, followed by a discussion of a procedure presented by Fair (1970) for obtaining consistent and asympotically efficient estimates of the structural parameters using a modified version of 2SLS (denoted, following Fair [1970] and Amemiya [1966], as *S2SLS*).

The identification question is at the heart of applying structural equations to XLPC designs; it is desired to not only exactly identify or overidentify the equations, but to include all major causal variables in the equations, particularly those providing spurious relationships (e.g., synchronous or cross-lagged common factors [cf. Kenny, 1973]). As discussed earlier, not all causal variables will likely be included, thus creating a specification error. However, an emphasis on including multiple sources of relevant causality provides a more explanatory theoretical network as well as an opportunity to test competing hypotheses in overidentified models.

For exemplary purposes, only enough exogenous variables were added to equations 17 and 18 to exactly identify the equations. This involved adding variables X_{1t} and X_{2t} to equation 17, and variables X_{2t} and X_{3t} to equation 18.[5] The exogenous variables were assumed to be independent of the disturbance terms in each equation. The new equations are (in deviation form)

$$y_{1t} = b_{12}\, y_{2t} + c_{11}\, x_{1t} + c_{12}\, x_{2t} + c_{13}\, y_{2t-1} + c_{14}\, y_{1t-1} + g_{1t} \tag{19}$$

$$y_{2t} = b_{21}\, y_{1t} + c_{22}\, x_{2t} + c_{23}\, x_{3t} + c_{24}\, y_{1t-1} + c_{25}\, y_{2t-1} + g_{2t} \tag{20}$$

where g_{1t} and g_{2t} are the disturbance terms; serial correlation between the disturbance terms is represented by $g_{1t} = p_{11}\, g_{1t-1} + h_{1t}$ and $g_{2t} = p_{22}\, g_{2t-1} + h_{2t}$; and all previous probability limits remain the same with d replaced by g, and ε replaced by h.

The application of S2SLS proceeds in the following manner. To conserve space, the equation for y_{1t} received focus. First, the generalized differencing process is applied to equation 19 in order to replace serially correlated g_{1t} with the random component h_{1t}.

$$\underline{y_{1t}} - \underline{\hat{p}_{11}}\ \underline{y_{1t-1}} = \underline{b_{12}}\ (\underline{y_{2t}} - \underline{\hat{p}_{11}}\ \underline{y_{2t-1}}) + \underline{c_{11}}\ (\underline{x_{1t}} - \underline{\hat{p}_{11}}\ \underline{x_{1t-1}})$$
$$+ \underline{c_{12}}\ (\underline{x_{2t}} - \underline{\hat{p}_{11}}\ \underline{x_{2t-1}}) + \underline{c_{13}}\ (\underline{y_{2t-1}} - \underline{\hat{p}_{11}}\ \underline{y_{2t-2}})$$
$$+ \underline{c_{14}}\ (\underline{y_{1t-1}} - \underline{\hat{p}_{11}}\ \underline{y_{1t-2}}) + [(\underline{p_{11}} - \underline{\hat{p}_{11}})\ \underline{g_{1t-1}} + \underline{h_{1t}}] \tag{21}$$

where $\underline{\hat{p}_{11}}$ is an estimate of $\underline{p_{11}}$, the disturbance term will equal $\underline{h_{1t}}$ if an appropriate least squares estimate of p_{11} is obtained (discussed later), and an <u>additional</u> <u>wave</u> of data must be collected for the endogenous and exogenous variables.

The second step involves the first-stage of S2SLS. To visualize the development of the reduced form, it should be noted that the only variable in equation 21 correlated with the new error term ($\underline{h_{1t}}$) in the limit is $\underline{y_{2t}}$, which is due to the nonrecursiveness of the equation. All other variables in equation 21 are predetermined (i.e., lagged endogenous, cross-lagged endogenous, and current and lagged exogenous), and are not correlated with ($\underline{h_{1t}}$) in the limit as a result of either assumptions or the generalized differencing process. Thus, a reduced form is needed to obtain a predicted score for $\underline{y_{2t}}$ based upon all predetermined variables in equation 21 <u>and</u> the predetermined variables that would be obtained from applying the generalized differencing process to equation 20.

In general terms, using reduced form estimated parameters, the reduced form for $\underline{y_{2t}}$ is

$$\underline{\hat{y}_{2t}} = \underline{\hat{\pi}_{21}}\ \underline{y_{2t-1}} + \underline{\hat{\pi}_{22}}\ \underline{y_{2t-2}} + \underline{\hat{\pi}_{23}}\ \underline{y_{1t-1}} + \underline{\hat{\pi}_{24}}\ \underline{y_{1t-2}} + \underline{\hat{\pi}_{25}}\ \underline{x_{1t}}$$
$$+ \underline{\hat{\pi}_{26}}\ \underline{x_{1t-1}} + \underline{\hat{\pi}_{27}}\ \underline{x_{2t}} + \underline{\hat{\pi}_{28}}\ \underline{x_{2t-1}} + \underline{\hat{\pi}_{29}}\ \underline{x_{3t}} + \underline{\hat{\pi}_{210}}\ \underline{x_{3t-1}} + \underline{m_{2t}}$$

where the predicted scores for y_{2t} are direct functions of the predetermined variables, and $m_{2t} = (b_{21}\ h_{1t} + h_{2t}) \ / \ (1/[1-b_{21}\ b_{12}])$. OLS provides the predicted y_{2t} scores.

The third step in the procedure is to replace $(y_{2t} - \hat{p}_{11}\ y_{2t-1})$ with $(\hat{y}_{2t} - \hat{p}_{11}\ y_{2t-1})$ in equation 21, where plim $(\frac{1}{n}\ \hat{y}_{2t}\ h_{1t})$ is now equal to zero. If $y_{2t} - \hat{y}_{2t}$ is set equal to $\hat{v}_{1t}$, the new disturbance term for equation 21 for the second-stage regression is equal to $[(p_{11} - \hat{p}_{11})\ g_{1t-1} + h_{1t} + b_{12}\ \hat{v}_{1t}]$. The second-stage of 2SLS is then conducted using OLS. However, because p_{11} can only be estimated, several OLS analyses are conducted, using values of $\hat{p}_{11}$ varying between 1 and -1 (or an iterative procedure is used). The OLS analysis with the estimated value of p_{11} which yields the smallest sum of squared residuals of the second-stage regression, and the corresponding estimates of the structural parameters, is selected as a solution (an iterative procedure is provided by Fair [1970, p. 509]).

The minimum sum of squared residuals occurs where $\hat{p}_{11}$ equals p_{11} (in large samples), leaving the error term $(h_{1t} + b_{12}\ \hat{v}_{12})$, which has a zero expected value and *in the limit* is uncorrelated with $\hat{y}_{2t}$ as well as with the predetermined variables. That is, the predetermined variables are neither correlated with h_{1t}, for reasons already discussed, nor with $b_{12}\ \hat{v}_{1t}$ because they were used as predictors in the first-stage of S2SLS. The values for $\hat{y}_{2t}$ are uncorrelated with the disturbance term in the second-stage of S2SLS based on the logic presented earlier for 2SLS. (Another reason for employing lagged endogenous, cross-lagged endogenous, and current and lagged exogenous variables in the first-stage regression is to insure the orthogonality of $\hat{v}_{1t}$ and g_{1t-1}, which is necessary if the minimum sum of squared residuals is

to occur where $\underline{\hat{p}_{11}}$ equals $\underline{p_{11}}$ [Fair, 1970, p. 509]). Thus, the S2SLS procedure provides consistent estimates of the structural parameters.

For tests of goodness of fit, one would have to begin with overidentified structural equations, and, using the logic presented for nonrecursive models, add predetermined variables to the equations until exact identification was achieved. The S2SLS procedure would then be repeated on the exactly identified equations and the resulting parameter estimates examined to ascertain if they diverged from the assumed causal model.[6]

In summary, the application of structural equations and S2SLS to the XLPC design, or more generally to models involving lagged endogenous variables and nonrecursive relationships, is a rather complex process, complexity being interpreted in terms of the theoretical system that is required, the assumptions that must be made, the amount of data that must be collected, and the statistical procedures that are necessitated. For example, at least three waves of data must be collected (which includes the exogenous variables if lagged values of such are included in the equations prior to differencing). The variables must be highly reliable for reasons of parameter estimation, the calculation of difference scores (cf. Cronbach & Furby, 1970; Lord & Novick, 1968), and the possibility that the measurement errors of unreliable variables could be serially correlated over time (cf. Namboodiri et al., 1975; Pindyck & Rubinfeld, 1976). Furthermore, the inclusion of lagged variables in the equations may well present a problem of multicollinearity (cf. Johnston, 1972). Thus, while multiple lags are desirable to analyze such possibilities as positive or negative feedback loops (cf. Miller, 1971; Pelz & Lew, 1970), the addition of lagged variables to the equations may be

dysfunctional from another standpoint. Finally, the time lags and the model in general must be correctly specified, both of which are sizeable requirements. For example, the model may not be first-order autoregressive as assumed here, nor may the stationarity assumption be viable (see Pindyck and Rubinfeld [1976] for time-series procedures that might be employed when these assumptions are violated). A related concern is differing stabilities of the causal factors over time, where for example differing degrees of stability (short-term versus long-term) have different effects on the magnitude and even the sign of the parameter estimates, especially if the measurement periods do not correspond to the causal intervals (cf. Fisher, 1971; Pelz & Lew, 1970).

Unfortunately, space does not permit further consideration of the above issues, and the reader is referred to the cited references for additional reading. This section is concluded by simply reiterating that the XLPC design, or some variation thereof, is perhaps the most likely candidate at the present time in psychology for analyzing designs which include lagged endogenous variables and recursive relationships. This should not be construed to mean that the XLPC is a nondemanding procedure. The truth of the matter is that XLPC designs are quite demanding with respect to assumptions and data, although not as demanding as structural equations. Nevertheless, the development of appropriate structural models, which may include nonrecursive relationships, is a desirable goal for psychology because they provide a stronger theoretical and causal foundation.

DISCUSSION

The major goal of the present report has been to introduce psychologists to the rationale, assumptions, and analytical procedures of 2SLS and its

applications to selected structural equations. These include nonrecursive structural equations, structural equations which contain predictors with random measurement error, and structural equations which involve lagged values of endogenous variables. In addition, nonrecursive relationships were included in the last application in order to increase generality and to demonstrate how two of the applications could be combined. The first and last applications, however, assumed perfectly reliable variables (or, from a pragmatic standpoint, highly reliable variables). If this assumption is not met, then a procedure such as outlined in the second application (i.e., the use of instrumental variables) could be added to the analysis, although the instrumental variable approach is not a panacea for variables with large measurement errors.

The selected applications were considered to be reflective of many psychological phenomena. In particular, nonrecursive models may add a new dimension of analysis to the current Zeitgeist of interactionism in psychology. Moreover, in the presence of strong theory, hopefully based in part on previous research, the use of cross-sectional data should not preclude the development of at least tentative causal models if in fact assumptions have been reasonably met. Such research can provide a strong foundation on which to proceed to dynamic models, which presumably provide a stronger test of the model. It is of utmost importance to note that a crucial issue in the use of dynamic models and lagged variables is the degree to which the measurement of variables corresponds to real-world temporal sequences and time lags. Furthermore, assumptions such as the equilibrium-type condition are maintained in the dynamic analyses. That is, the stationarity assumption

of cross-lagged panel correlation requires that the causal process is in equilibrium (i.e., the structural equation for a variable is invariant with respect to time of measurement). However, if the assumptions for dynamic analysis are met, opportunities are provided to address and to test directly several issues which are typically assumptions and not wholly testable in cross-sectional analyses. These issues include the source and direction of causation, the necessity-sufficiency of causation, and dynamic-static causal relationships (cf. Feldman, 1975). On the other hand, if assumptions regarding dynamic analyses cannot be met, analyses based on cross-sectional data _may_ provide the more meaningful results, particularly if an equilibrium-type condition exists and the unmet assumption for dynamic analysis is measurement corresponding to real-world temporal sequences.

In conclusion, causal analyses that employ structural equations _and_ passive data, either lagged or nonlagged, have as a primary focus the identification of untenable causal models rather than the identification of a "true" causal model. This requires the use of overidentified models so that different models may be tested, and emphasis is placed on conceptual issues, rationale, and assumptions, and the internal consistency of results with respect to the theory, rationale, and assumptions. The strength of these models lies in nonexperimental inference (Kenny, 1975), even though problems concerning causal inferences associated with passive data obtained from natural observation and thus lacking randomization and control are well known.

Reference Notes

1. King, D. Feedback and reciprocal causation: A Monte Carlo study of the relative small sample properties of simultaneous equation estimators. Unpublished dissertation, University of North Carolina, 1974.

References

Amemiya, T. Specification analysis in the estimation of parameters of a simultaneous equation model with autoregressive residuals. Econometrica, 1966, 34, 283-306.

Barrow, J. C. Worker performance and task complexity as causal determinants of leader behavior style and flexibility. Journal of Applied Psychology, 1976, 61, 433-440.

Baseman, R. L. A generalized classical model of linear estimation of coefficients in a structural equation. Econometrica, 1957, 25, 77-83.

Blalock, H. M. Theory construction. Englewood Cliff, N. J.: Prentice-Hall, 1969.

Blalock, H. M. Estimating measurement error using multiple indicators and several points in time. American Sociological Review, 1970, 35, 101-111.

Blalock, H. M., Wells, C. S., & Carter, L. F. Statistical estimation with random measurement error. In E. F. Borgatta & G. W. Bohrnstedt (Eds.), Sociological methodology. San Francisco: Jossey-Bass, 1970.

Bohrnstedt, G. W. Observations on the measurement of change. In E. F. Borgatta & C. W. Bohrnstedt (Eds.), Sociological methodology. San Francisco: Jossey-Bass, 1969.

Bohrnstedt, G. W., & Carter, T. M. Robustness in regression analysis. In H. L. Costner (Ed.), Sociological methodology. San Francisco: Jossey-Bass, 1971.

Bowers, K. S. Situationism in psychology: An analysis and a critique. Psychological Review, 1973, 80, 307-336.

Boyle, R. P. Path analysis and ordinal data. American Journal of Sociology, 1970, 75, 461-480.

Campbell, D. T. From description to experimentation: Interpreting trends as quasi-experiments. In C. W. Harris (Ed.), Problems in measuring change. Madison: University of Wisconsin Press, 1963.

Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research in teaching. In N. L. Gage (Ed.), Handbook of research on teaching. New York: Rand McNally, 1963.

Christ, C. Econometric models and methods. New York: Wiley, 1966.

Cook, T. D., & Campbell, D. T. The design and conduct of quasi-experiments and true experiments in field settings. In M. D. Dunnette (Ed.), Handbook of industrial and organizational psychology. Chicago: Rand McNally, 1976.

Costner, H. L. Theory, deduction, and rules of correspondence. American Journal of Sociology, 1969, 75, 245-263.

Costner, H. L., & Schoenberg, R. Diagnosing indicator ills in multiple indicator models. In A. S. Goldberger & O. D. Duncan (Eds.), Structural equation models in the social sciences. New York: Seminar Press, 1973.

Cronbach, L. J., & Furby, L. How should we measure change - or should we? Psychological Bulletin, 1970, 74, 68-80.

Cummins, R. C. Leader-member relations as a moderator of the effects of leader behavior and attitude. Personnel Psychology, 1972, 25, 655-660.

Dansereau, F., Jr., Graen, G., & Haga, W. J. A vertical dyad linkage approach to leadership within formal organizations, a longitudinal investigation of the role making process. Organizational Behavior and Human Performance, 1975, 13, 46-78.

Darlington, R. B. Multiple regression in psychological research and practice. Psychological Bulletin, 1968, 69, 161-182.

Darlington, R. B., & Rom, J. F. Assessing the importance of independent variables in nonlinear causal laws. American Educational Research Journal, 1972, 9 pages.

Duncan, O. D. Path analysis: Sociological examples. American Journal of Sociology, 1966, 72, 1-16.

Duncan, O. D. Some linear models for two-wave, two-variable panel analysis. Psychological Bulletin, 1969, 72, 177-182.

Duncan, O. D. Duncan's corrections of published text of "Peer influences on aspirations: A reinterpretation." American Journal of Sociology, 1970, 75, 1042-1046.

Duncan, O. D. Introduction to structural equation models. New York: Academic Press, 1975.

Duncan, O. D., Haller, A. O., & Portes, A. Peer influences on aspirations: A reinterpretation. American Journal of Sociology, 1968, 74, 119-137.

Duncan, O. D., Haller, A. O., & Portes, A. Peer influences on aspirations: A reinterpretation. In H. M. Blalock (Ed.), Causal models in the social sciences. New York: Aldine-Atherton, 1971.

Ekehammar, B. Interactionism in personality from a historical perspective. Psychological Bulletin, 1974, 81, 1026-1048.

Endler, N. S. The case for person-situation interactions. Canadian Psychological Review, 1975, 16, 12-21.

Endler, N. S., & Magnusson, D. Toward an interactional psychology of personality. Psychological Bulletin, 1976, 83, 956-974.

Evans, M. G. A leader's ability to differentiate, the subordinate's perception of the leader, and the subordinate's performance. Personnel Psychology, 1973, 26, 385-395.

Fair, R. C. The estimation of simultaneous equation models with lagged endogenous variables and first order serially correlated errors. Econometrica, 1970, 38, 507-516.

Farris, F., & Lim, F. Effects of performance on leadership cohesiveness, influence, satisfaction, and subsequent performance. Journal of Applied Psychology, 1969, 53, 490-497.

Feldman, J. Considerations in the use of causal-correlational techniques in applied psychology. Journal of Applied Psychology, 1975, 60, 663-670.

Fiedler, F. E., & Chemers, M. M. Leadership and effective management. New York: Scott-Foresman, 1974.

Fisher, F. M. The identification problem in econometrics. New York: McGraw-Hill, 1966.

Fisher, F. M. The choice of instrumental variables in the estimation of economy-wide econometric models. In H. M. Blalock, Jr. (Ed.), Causal models in the social sciences. Chicago: Aldine-Atherton, 1971.

Gibb, C. Leadership. In G. Lindzey & E. Aronson (Eds.), The handbook of social psychology, Vol. 4. Massachusetts: Addison-Wesley, 1969.

Goldberger, A. S. Discerning a causal pattern among data on voting behavior. In H. M. Blalock, Jr. (Ed.), Causal models in the social sciences. Chicago: Aldine-Atherton, 1971.

Goldberger, A. S. Efficient estimation in overidentified models: An interpretive analysis. In A. S. Goldberger & O. D. Duncan (Eds.), Structural equation models in the social sciences. New York: Seminar Press, 1973.

Goldberger, A. S., & Duncan, O. D. Structural equation models in the social sciences. New York: Seminar Press, 1973.

Greene, C. N. Causal connections among managers merit pay, satisfaction, and performance. Journal of Applied Psychology, 1973, 58, 95-100.

Greene, C. N. The reciprocal nature of influence between leader and subordinate. Journal of Applied Psychology, 1975, 60, 187-193.

Guion, R. M. Recruiting, selection, and job placement. In M. D. Dunnette (Ed.), Handbook of industrial and organizational psychology. Chicago: Rand McNally, 1976.

Hauser, R. M., & Goldberger, A. S. The treatment of unobservable variables in path analysis. In H. L. Costner (Ed.), Sociological methodology, 1971. San Francisco: Jossey-Bass, 1971.

Heise, D. R. Causal inference from panel data. In E. F. Borgatta and G. W. Bohrnstedt (Eds.), Sociological methodology. San Francisco: Jossey-Bass, 1970.

Heise, D. R. Causal analysis. New York: John Wiley & Sons, 1975.

Herold, D. M. Interaction of subordinate and leader characteristics in moderating the consideration-satisfaction relationship. Journal of Applied Psychology, 1974, 59, 649-651.

Hill, W. A., & Hughes, D. Variations in leader behavior as a function of task type. Organizational Behavior and Human Performance, 1974, 11, 83-96.

House, R. J., & Mitchell, T. R. Path-goal theory of leadership. Journal of Contemporary Business, Autumn, 1974, 81-97.

Indik, B. P. The scope of the problem and some suggestions toward a solution. In B. P. Indik & F. K. Berrien (Eds.), People, groups and organizations. New York: Teachers College Press, 1968.

James, L. R., & Jones, A. P. Organizational structure: A review of structural dimensions and their conceptual relationships with attitudes and behavior. Organizational Behavior and Human Performance, 1976, 16, 74-113.

Johnston, J. J. Econometric methods (2nd ed.). New York: McGraw-Hill, 1972.

Jöreskog, K. G. A general method for the analysis of covariance structures. Biometrika, 1970, 57, 239-251.

Jöreskog, K. G. A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), Structural equation models in the social sciences. New York: Seminar Press, 1973.

Kalleberg, A. L., & Kluegel, J. R. Analysis of the multitrait-multimethod matrix: Some limitations and an alternative. Journal of Applied Psychology, 1975, 60, 1-9.

Katz, D., & Kahn, R. L. The social psychology of organizations. New York: Wiley, 1966.

Kenny, D. A. Cross-lagged and synchronous common factors in panel data. In A. S. Goldberger & O. D. Duncan (Eds.), Structural equation models in the social sciences. New York: Seminar Press, 1973.

Kenny, D. A. Cross-lagged panel correlation: A test for spuriousness. Psychological Bulletin, 1975, 82, 887-903.

Kerlinger, F. N., & Pedhazur, E. J. Multiple regression in behavioral research. New York: Holt, Rinehart and Winston, 1973.

Kerr, S., & Schriesheim, C. Consideration, initiating structure, and organizational criteria - an update of Korman's 1966 review. Personnel Psychology, 1974, 27, 555-568.

Kohn, M. L., & Schooler, C. Occupational experience and psychological functioning: An assessment of reciprocal effects. American Sociological Review, 1973, 38, 97-118.

Lichtman, C. M., & Hunt, R. G. Personality and organization theory: A review of some conceptual literature. Psychological Bulletin, 1971, 76, 271-294.

Likert, R. The human organization. New York: McGraw-Hill, 1967.

Lord, F. M., & Novick, M. R. Statistical theory of mental test scores. Reading: Addison-Wesley, 1968.

Lowin, A., & Craig, J. R. The influence of level of performance on managerial style: An experimental object-lesson in the ambiguity of correlational data. Organizational Behavior and Human Performance, 1968, 3, 440-458.

Mason, R., & Halter, A. N. The application of a system of simultaneous equations to an innovation diffusion model. Social Forces, 1968, 47, 182-195.

Miller, A. D. Logic of causal analysis: From experimental to nonexperimental designs. In H. M. Blalock, Jr. (Ed.), Causal models in the social sciences. Chicago: Aldine-Atherton, 1971.

Namboodiri, N. K., Carter, L. R., & Blalock, H. M., Jr. Applied multivariate analysis and experimental designs. New York: McGraw-Hill, 1975.

Nerlove, M. Further evidence on the estimation of dynamic economic relations from a time series of cross sections. Econometrica, 1971, 39, 359-382.

Oldham, G. R. The motivational strategies used by supervisors: Relationships to effectiveness indicators. Organizational Behavior and Human Performance, 1976, 15, 66-86.

Overton, W. F., & Reese, H. W. Models of development: Methodological implications. In J. R. Nesselroade & H. W. Reese (Eds.), Life-span developmental psychology: Methodological issues. New York: Academic Press, 1973.

Parker, W. S., Jr. Black-white differences in leader behavior related to subordinates' reactions. Journal of Applied Psychology, 1976, 61, 140-147.

Pelz, D. C., & Lew, R. A. Heise's causal model applied. In E. F. Borgatta & G. W. Bohrnstedt (Eds.), Sociological methodology. San Francisco: Jossey-Bass, 1970.

Pindyck, R. S., & Rubinfeld, D. L. Econometric models and economic forecasts. New York: McGraw-Hill, 1976.

Saunders, D. R. Moderator variables in prediction. Educational and Psychological Measurement, 1956, 16, 209-222.

Sells, S. B. An interactionist looks at the environment. *American Psychologist*, 1963, *18*, 696-702.

Sells, S. B. General theoretical problems related to organizational taxonomy: A model solution. In B. P. Indik and F. K. Berrien (Eds.), *People, groups and organizations*. New York: Teachers College Press, 1968.

Sims, H. P., & Szilagyi, A. D. Leader structure and subordinate satisfaction for two hospital administrative levels: A path analysis approach. *Journal of Applied Psychology*, 1975, *60*, 194-197.

Singh, B. K. Path analysis in social science research. In B. K. Singh (Ed.), "Recent trends in social analysis", *Indian Journal of Extension Education: Special Volume on Methodology*, 1975.

Singh, B. K., & Williams, J. S. Feedback in path analysis: Some sociological examples. *Social Statistics Proceedings*, American Statistical Association, 1972, 426-431.

Sockloff, A. L. The analysis of nonlinearity via linear regression with polynominal and product variables: An examination. *Review of Educational Research*, 1976, *46*, 267-291.

Spaeth, J. L. Path analysis. In D. J. Amick & H. J. Walberg (Eds.), *Introductory multivariate analysis for educational, psychological, and social research*. Berkeley: McCutchan, 1975.

Steers, R. M. Task-goal attributes, *n* achievement, and supervisory performance. *Organizational Behavior and Human Performance*, 1975, *13*, 392-403.

Stogdill, R. M. *Handbook of leadership*. Glencoe, Ill.: Free Press, 1974.

Strotz, R. H., & Wold, H. O. Recursive versus nonrecursive systems: An attempt at synthesis. In H. M. Blalock, Jr. (Ed.), Causal models in the social sciences. Chicago: Aldine-Atherton, 1971.

Theil, H. Repeated least-squares applied to complete equation systems. The Hague: Central Planning Bureau, 1953 (mimeographed). (a)

Theil, H. Estimation and simultaneous correlation in complete equation systems. The Hague: Central Planning Bureau, 1953 (mimeographed). (b)

Theil, H. Principles of econometrics. New York: John Wiley, 1971.

Vroom, V. H. Leadership. In M. D. Dunnette (Ed.), Handbook of industrial and organizational psychology. Chicago: Rand McNally, 1976.

Waite, L. J., & Stolzenberg, R. M. Intended childbearing and labor force participation of young women: Insights from nonrecursive models. American Sociological Review, 1976, 41, 235-252.

Wallis, K. F. Lagged dependent variables and serially correlated errors: A reappraisal of three-pass least squares. Review of Economics and Statistics, 1967, 49, 555-567.

Werts, C. E., & Linn, R. L. Path analysis: Psychological examples. Psychological Bulletin, 1970, 74, 193-212.

Werts, C. E., Linn, R. L., & Jöreskog, K. G. Estimating the parameters of path models involving unmeasured variables. In H. M. Blalock, Jr. (Ed.), Causal models in the social sciences. Chicago: Aldine-Atherton, 1971.

Wiley, D. E. The identification problem for structural equation models with unmeasured variables. In A. S. Goldberger & O. D. Duncan (Eds.), Structural equation models in the social sciences. New York: Seminar Press, 1973.

Wiley, D. E., & Wiley, J. A. The estimation of measurement error in panel data. In H. M. Blalock, Jr. (Ed.), Causal models in the social sciences. Chicago: Aldine-Atherton, 1971.

Wold, H. A. Path models with latent variables that have several indicators: Partings of the ways in the light of NIPALS modeling. In H. M. Blalock, A. Aganbegian, F. Borodkin, R. Boudon, & V. Capecchi (Eds.), Quantitative sociology. New York: Academic Press, 1975.

Wright, S. The method of path coefficients. Annals of Mathematical Statistics, 1934, 5, 161-215.

Footnotes

Support for this project was provided under Office of Naval Research Contract N00014-76-C-0008, Office of Naval Research Project RR042-08-01 N170-743, and by the Bureau of Medicine and Surgery, Department of the Navy, Under Research Work Unit 51 002-5015 DXSF. Opinions expressed are those of the authors and are not to be construed as necessarily reflecting the official view or endorsement of the Department of the Navy.

The authors wish to thank S. B. Sells, R. G. Demaree, G. W. Joe, G. A. Schieren, E. A. Hartman, J. R. Bruni, and C. H. Hornick for their helpful suggestions and advice.

Request for reprints should be sent to Lawrence R. James, Institute of Behavioral Research, Texas Christian University, Fort Worth, Texas 76129.

[1]The equations presented in this paper employ unstandardized variables and unstandardized regression weights as estimates of (unstandardized) structural parameters. The use of standardized variables and their corresponding "beta" weights has been quite popular because structural equations can then be addressed in a "path analysis" paradigm. However, we have chosen to employ unstandardized variables because standardization of variables might obscure distinctions between estimates of structural parameters and the variances-covariances that describe joint distributions of variables in a population. That is, if it is possible for a variable to have different distributions for reasons such as the use of different populations or changes in a particular population over time, then unstandardized regression weights should be employed because they are still comparable across the distributions while standardized regression weights are not (cf. Namboodiri et al., 1975; Wiley & Wiley, 1971).

Footnotes (cont'd)

[2]The authors would like to thank an unknown reviewer for pointing out that a sufficient condition for equilibrium would be "to allow some movements of the dependent variable values if these are compensated by the inverse movement of the dependent variables by other individuals with similar values on the exogenous variables".

[3]The total number of alternative causal models that might be tested in many designs could be considered infinite. In this sample, we have only considered the causal relationships among observables, with the assumptions associated with nonrecursive models (e.g., correlated disturbances). However, the introduction of nonobservables, the possibility serially correlated disturbances, and so forth could greatly extend the complexity of the model as well as competing causal hypotheses. On the other hand, the number of competing causal hypotheses is usually reduced depending on the causal closure and the nature of theoretical orientation. That is, if there is a compelling reason to believe that there are only a few causal structures that would be meaningful, there is no reason to test all possible permutations and combinations which may emerge. For example, if one is fairly sure of the fact that there is an asymmetric causal relationship which will hold, the number of alternative models are automatically reduced and would not require further testing.

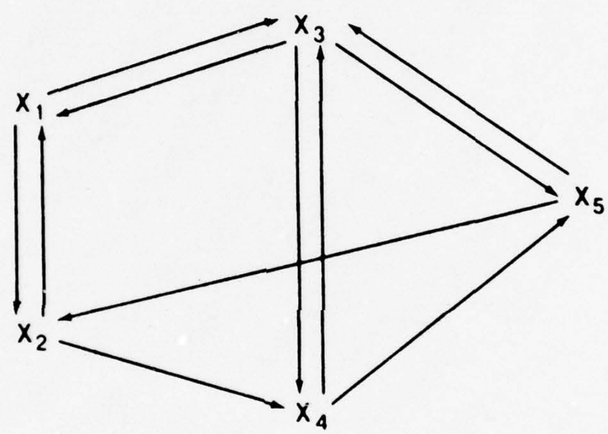
[4]There is, of course, the possibility of recursive relationships, which include the leader causing subordinate behaviors, the subordinate causing leader behaviors, and various feedback loops with known time lags. To avoid complexity, we have not addressed these possibilities here.

Footnotes (cont'd)

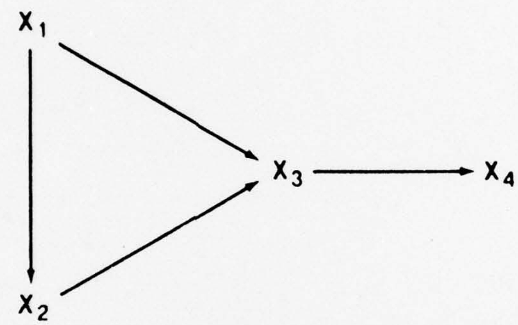
[5]The additional inclusion of lagged values of the exogenous variables would present a more realistic example; however, we are attempting to provide a general introduction and to minimize complexity. On the other hand, as will be shown, lagged values of the exogenous variables will enter into the analysis.

[6]As noted earlier, one reason for employing a nonrecursive model in the present discussion was that a reciprocal relationship between the dependent variables provided a competing hypothesis for the XLPC design, which traditionally has been viewed as an asymmetric causal model. However, the S2SLS and goodness of fit tests may indicate that the structural model is recursive rather than nonrecursive (i.e., the nonrecursive relationships are not empirically substantiated). In this circumstance, or in cases where nonrecursive relationships can be ruled out a priori, a different application of 2SLS may be employed when the XLPC design is viewed in terms of structural equations (or in more general terms, when the model is recursive and includes lagged endogenous variables with serially correlated disturbances). As described by Johnston (1972) and Wallis (1967), this procedure involves replacing the lagged endogenous variables with instruments, based on estimates provided by lagged values of exogenous variables, and then applying a generalized least squares estimation procedure to compute the second-stage regressions.

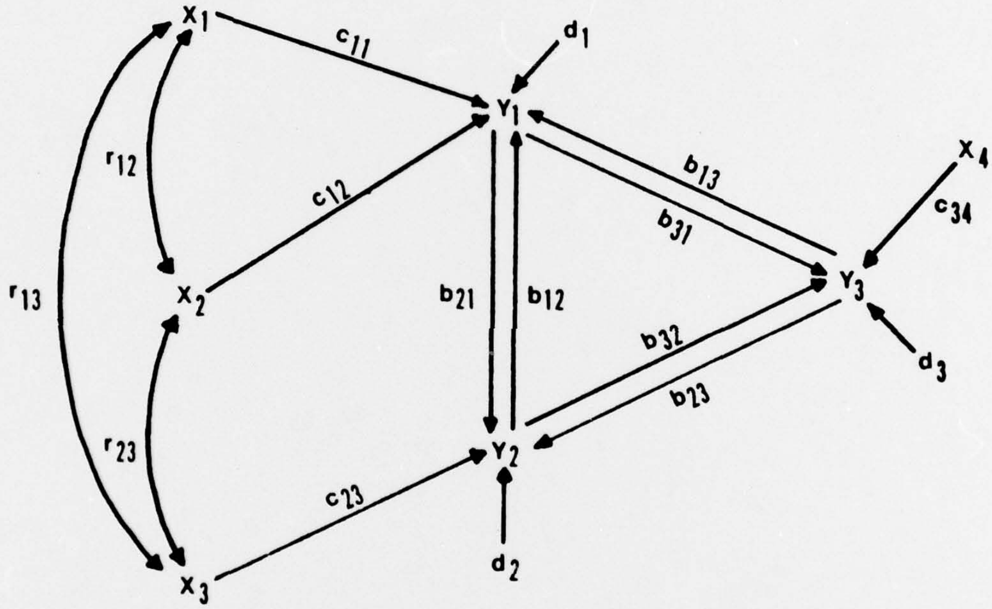
Figure Captions



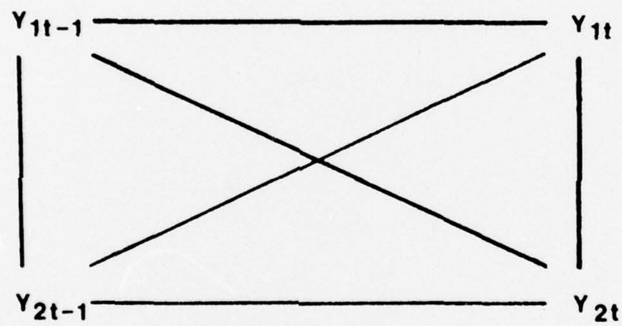
X_1
X_2
X_3
X_4
X_5

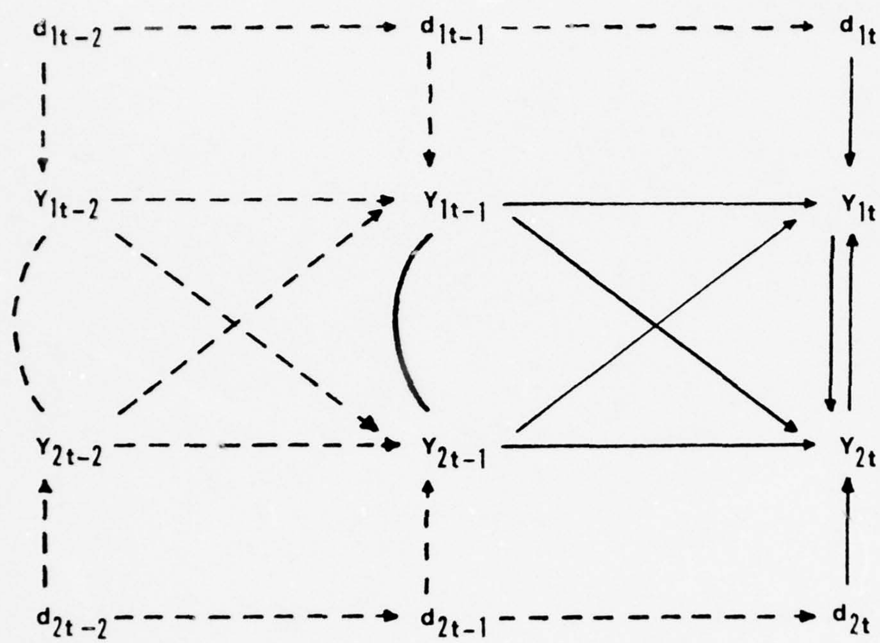


X1
X2
X3
X4



X_1
X_2
X_3
X_4
Y_1
Y_2
Y_3
c_{11}
c_{12}
c_{23}
c_{34}
r_{12}
r_{13}
r_{23}
b_{12}
b_{21}
b_{13}
b_{31}
b_{23}
b_{32}
d_1
d_2
d_3


Y_{1t-1}
Y_{1t}
Y_{2t-1}
Y_{2t}



d_{1t-2}
d_{1t-1}
d_{1t}
Y_{1t-2}
Y_{1t-1}
Y_{1t}
Y_{2t-2}
Y_{2t-1}
Y_{2t}
d_{2t-2}
d_{2t-1}
d_{2t}

Appendix A: Development of a Reduced Form in 2SLS

The reduced from for a set of equations is obtained by a process of substitution in which the endogenous variables appearing in a particular equation as predictors are replaced by the right-side of their respective structural equation. The solution for the value of each endogenous variable is then obtained in terms of the exogenous (predetermined) variables only, plus a disturbance term. For example, in the exactly identified, nonrecursive equations

$$\underline{y_1} = \underline{b_{12}}\ \underline{y_2} + \underline{c_{11}}\ \underline{x_1} + \underline{d_1} \qquad \text{(A.1)}$$

$$\underline{y_2} = \underline{b_{21}}\ \underline{y_1} + \underline{c_{22}}\ \underline{x_2} + \underline{d_2} \qquad \text{(A.2)}$$

$\underline{y_1}$ may be expressed as a function of the predetermined variables ($\underline{x_1}$ and $\underline{x_2}$) and a disturbance term by substituting the right-side of equation A.2 in equation A.1 to replace $\underline{y_2}$. The reduced form for $\underline{y_1}$ is therefore

$$\begin{aligned}\underline{y_1} &= \underline{b_{12}}\ (\underline{b_{21}}\ \underline{y_1} + \underline{c_{22}}\ \underline{x_2} + \underline{d_2}) + \underline{c_{11}}\ \underline{x_1} + \underline{d_1} \\ &= \underline{b_{12}}\ \underline{b_{21}}\ \underline{y_1} + \underline{b_{12}}\ \underline{c_{22}}\ \underline{x_2} + \underline{b_{12}}\ \underline{d_2} + \underline{c_{11}}\ \underline{x_1} + \underline{d_1} \\ &= \frac{1}{(1-\underline{b_{12}}\ \underline{b_{21}})}\ [\underline{b_{12}}\ \underline{c_{22}}\ \underline{x_2} + \underline{c_{11}}\ \underline{x_1} + (\underline{d_1} + \underline{b_{12}}\ \underline{d_2})] \qquad \text{(A.3)}\end{aligned}$$

In more general terms, the reduced form for the $\underline{y_1}$ and $\underline{y_2}$ equations is viewed as

$$\underline{\hat{y}_1} = \underline{\hat{\pi}_{11}}\ \underline{x_1} + \underline{\hat{\pi}_{12}}\ \underline{x_2} + \underline{m_1} \qquad \text{(A.4)}$$

$$\underline{\hat{y}_2} = \underline{\hat{\pi}_{21}}\ \underline{x_1} + \underline{\hat{\pi}_{22}}\ \underline{x_2} + \underline{m_2} \qquad \text{(A.5)}$$

where $\hat{y}_1$ and $\hat{y}_2$ represent the predicted values of the endogenous variables, $\hat{\pi}_{gk}$ represent the *unbiased* estimates of the population reduced form parameters (π_{gk}), and the m_g represent the disturbance terms for the reduced form.

The predicted values $\hat{y}_1$ and $\hat{y}_2$ are obtained by simply applying OLS to equations A.4 and A.5, which also provides the estimates $\hat{\pi}_{gk}$. Of interest, however, is the fact that the (estimated) reduced form parameters are exact nonlinear functions of the (estimated) structural parameters, and vice-versa (Duncan, 1975). For example, $\hat{\pi}_{11}$ in equation A.4 is equal to $(b_{12}\ c_{22}\ x_2)$ / $(1/[1-b_{12}\ b_{21}])$.

Appendix B: Matrix Algebra for Applying 2SLS to Nonrecursive Structural Equations and the Rank Condition

As presented by Johnston (1972), a particular equation selected from a set of simultaneous, nonrecursive equations may be viewed as

$$\mathbf{y} = \mathbf{Y}_1 \mathbf{b} + \mathbf{X}_1 \mathbf{c} + \mathbf{d} \qquad \text{(B.1)}$$

where

$\mathbf{y}$ is an n x 1 vector of observations (raw scores) on the dependent variable,

$\mathbf{Y}_1$ is an n x g matrix of observations on mutually interacting endogenous variables included in the equation,

$\mathbf{b}$ is a g x 1 vector of estimated structural parameters attached to the $\mathbf{Y}_1$ variables,

X_1 is an $n \times k$ matrix of observations on the exogenous (predetermined) variables in the equation (a column of ones is included if an intercept is required),

c is a $k \times 1$ vector of estimated structural parameters for the X_1 variables, and

d is an $n \times 1$ vector of disturbances for this equation.

It is assumed that

$$\text{plim } (\tfrac{1}{n} X_1' d) = 0, \text{ while plim } (\tfrac{1}{n} Y_1' d) \neq 0$$

The first stage regression or reduced form is

$$\hat{Y}_1 = X (X' X)^{-1} X' Y_1 \qquad \text{(B.2)}$$

where $X = [X_1 \; X_2]$, which is equal to the $n \times k$ matrix of observations on all exogenous (predetermined) variables, given that X_2 is the matrix of observations on the exogenous (predetermined) variables not included in the equation under study.

The second-stage of 2SLS involves regressing y on $\hat{Y}_1$ (which replaces Y_1) and X_1. The estimating equations for this analysis are

$$\begin{bmatrix} \hat{Y}_1' \hat{Y}_1 & \hat{Y}_1' X_1 \\ X_1' \hat{Y}_1 & X_1' X_1 \end{bmatrix} \begin{bmatrix} \tilde{b} \\ c \end{bmatrix} = \begin{bmatrix} \hat{Y}_1' y \\ X_1' y \end{bmatrix} \qquad \text{(B.3)}$$

where $\tilde{b}$ and c are estimates of the population structural parameters based on 2SLS.

The necessary and sufficient rank condition for identification can be shown by designating X_2 as a column vector of exogenous variables excluded from equation B.1 (for derivation purposes, the references to observations

will be deleted, and population values are assumed). The equations for the reduced form corresponding to the $\mathbf{Y}_1$ variables only are then (Fisher, 1966, p. 52)

$$\hat{\mathbf{Y}}_1 = \mathbf{\Pi}^{11} \mathbf{X}_1 + \mathbf{\Pi}^{12} \mathbf{X}_2 + \mathbf{v}_1$$

where $\mathbf{X}_1$ is a column vector of exogenous variables included in the equation, $\mathbf{\Pi}^{11}$ and $\mathbf{\Pi}^{12}$ represent (population) reduced form parameters, and $\mathbf{v}_1$ is a column vector of reduced form disturbances.

A necessary and sufficient condition for identification of equation B.1 is that the rank of $\mathbf{\Pi}^{12}$ be equal to be number ($\underline{r}$) of endogenous variables included in $\mathbf{Y}_1$ ($\mathbf{\Pi}^{12}$ has $\underline{r}$ rows and $\underline{K} - \underline{f}$ columns, where $\underline{f}$ is equal to the number of exogenous variables in $\mathbf{X}_1$) (Fisher, 1966, p. 54).

Distribution List

Mandatory

Office of Naval Research (3 copies)
(Code 452)
800 N. Quincy St.
Arlington, VA 22217

Director
U. S. Naval Research Laboratory (6 copies)
Washington, DC 20390
ATTN: Technical Information Division

Defense Documentation Center (12 copies)
Building 5
Cameron Station
Alexandria, VA 22314

Library, Code 2029 (6 copies)
U. S. Naval Research Laboratory
Washington, DC 20390

Science & Technology Division
Library of Congress
Washington, DC 20540

ONR FIELD

Director
ONR Branch Office
536 S. Clark St.
Chicago, IL 60605

Research Psychologist
ONR Branch Office
536 S. Clark St.
Chicago, IL 60605

Principal Investigators

Dr. Alvin J. Abrams
Navy Personnel R & D Center
San Diego, CA 92152

Dr. Clayton P. Alderfer
Dept. of Administrative Sciences
Yale University
New Haven, CT 06520

Dr. James A. Bayton
Dept. of Psychology
Howard University
Washington, DC 20001

Dr. Carl Bennett
Battelle Memorial Institute
4000 N. E. 41st St.
Seattle, WA 98105

Dr. H. Russell Bernard
Dept. of Sociology & Anthropology
West Virginia University
Morgantown, WV 26506

Dr. Milton R. Blood
School of Business
Georgia Institute of Technology
Atlanta, GA 30332

Dr. Davis B. Bobrow
University of Maryland
College Park, MD 20742

Dr. David G. Bowers
Institute for Social Research
University of Michigan
Ann Arbor, MI 48106

Dr. Robyn M. Dawes
Oregon Research Institute
488 E. 11th Ave.
Eugene, OR 97403

Dr. Harry R. Day
University City Science Center
Center for Social Development
3508 Science Center
Philadelphia, PA 19104

Principal Investigators (continued)

Dr. Dynes
Ohio State University Research Foundation
1314 Kinnear Road
Columbus, OH 43212

Dr. Fred E. Fiedler
Department of Psychology
University of Washington
Seattle, WA 98105

Dr. Allan H. Fisher, Jr.
Hay Associates
1625 Eye St., N. W. Suite 1001
Washington, DC 20006

Dr. Samuel L. Gaertner
Department of Psychology
University of Delaware
220 Wolf Hall
Newark, DE 19711

Dr. Paul S. Goodman
Graduate School of Industrial Administration
Carneige-Mellon University, Schenley Park
Pittsburgh, PA 15213

Dr. Gloria L. Grace
System Development Corporation
2500 Colorado Ave.
Santa Monica, CA 90406

Dr. Eric Gunderson
Naval Health Research Center
San Diego, CA 92152

Dr. Richard Hackman
Department of Administrative Sciences
Yale University
New Haven, CT 06520

Dr. Thomas W. Harrell
Graduate School of Business
Stanford University
Stanford, CA 94305

Dr. Charles F. Hermann
Ohio State University Research Foundation
1314 Kinnear Road
Columbus, OH 43212

Dr. Charles L. Hulin
Department of Psychology
University of Illinois
Champaign, IL 61820

Dr. Norman J. Johnson
School of Urban & Public Affairs
Carnegie-Mellon University
Pittsburgh, PA 15213

Dr. David C. McClelland
McBer and Company
137 Newbury St.
Boston, MA 02139

Dr. Elliott M. McGinnies
Psychology Department
American University
Washington, DC 20016

Dr. Terence R. Mitchell
School of Business Administration
University of Washington
Seattle, WA 98195

Dr. Peter R. Monge
Department of Speech-Communication
California State University
San Jose, CA 95192

Dr. James A. Moore
CACI, Inc.
800 Garden St.
Santa Barbara, CA 93101

Dr. Stanley M. Nealey
Battelle Memorial Institute
4000 N. E. 41st St.
Seattle, WA 98105

Dr. Herbert R. Northrup
Industrial Research Unit
University of Pennsylvania
Philadelphia, PA 19174

Dr. Benson E. Penick
Carnegie-Mellon University
Margaret Morrison 410
Pittsburgh, PA 15213

Dr. Chester M. Pierce
Harvard University
Nichols House
Appian Way
Cambridge, MA 92138

Dr. Diane M. Ramsey-Klee
R-K Research & System Design
3947 Ridgemont Dr.
Malibu, CA 90265

Principal Investigators (continued)

Dr. Karlene H. Roberts
School of Business Administration
University of California
Berkeley, CA 94720

Dr. Moshe F. Rubinstein
University of California
405 Hilgard Ave.
Los Angeles, CA 90024

Dr. John Ruhe
University of North Carolina
Dept. of Business Administration
Charlotte, NC 28223

Dr. Rudolph J. Rummel
Political Science Department
University of Hawaii
Honolulu, HI 96822

Dr. Irwin Sarason
Department of Psychology
University of Washington
Seattle, WA 98195

Dr. Edgar H. Schein
Sloan School of Management
Massachusetts Institute of Technology
Cambridge, MA 02139

Dr. Saul B. Sells
Texas Christian University
Fort Worth, TX 76129

Dr. Siegfried Streufert
Department of Psychology
Purdue University
Lafayette, IN 47907

Dr. Richard E. Sykes
Minnesota Systems Research, Inc.
2412 University Ave., S. E.
Minneapolis, MN 55414

Dr. Lorand B. Szalay
American Institutes for Research
Foxhall Square
3301 New Mexico Ave., N. W.
Washington, DC 20016

Dr. H. H. Vreeland III
Human Sciences Research, Inc.
Westgate Research Park
7710 Old Springhouse Road
McLean, VA 22101

Dr. Victor H. Vroom
School of Organization & Management
Yale University
56 Hillhouse Ave.
New Haven, CT 06520

Dr. Paul Wall
Div. of Behavioral Science Research
Tuskegee Institute
Tuskegee, AL 36088

Dr. Wilkenfeld
University of Maryland
College Park, MD 20742

Dr. Philip G. Zimbardo
Dept. of Psychology
Stanford University
Stanford, CA 94305

MISCELLANEOUS

Air Force

AFOSR (NL)
1400 Wilson Blvd.
Arlington, VA 22209

Army

Army Research Institute (2 copies)
Commonwealth Bldg.
1300 Wilson Blvd.
Rosslyn, VA 22209

Coast Guard

Chief, Psychological Research Branch
U. S. Coast Guard (G-P-1/62)
400 7th St., S. W.
Washington, DC 20590

Marine Corps

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps
(Code RD-1)
Washington, DC 20380

MISCELLANEOUS (continued)

Navy

Chief of Naval Personnel
Assistant for Research Liaison (Pers-Or)
Washington, DC 20370

Bureau of Naval Personnel (Pers-6)
Assistant Chief of Naval Personnel for
Human Goals
Washington, DC 20370

Cdr. Paul D. Nelson, MSC, USN
Head, Human Performance Division (Code 44)
Navy Medical R & D Command
Bethesda, MD 20014

LCdr. C. A. Patin, U. S. N.
Director, Human Goals Dept.
Code 70, Naval Training Center
Orlando, FL 32813

Office of Manpower Management
Personnel Management Evaluation Branch (72)
Washington, DC 20390

Assistant Officer in Charge
Naval Internal Relations Activity
Pentagon, Room 2E329
Washington, DC 20350

Naval Postgraduate School
Monterey, CA 93940
ATTN: Library (Code 2124)

Professor John Senger
Operations Research &
Administration Sciences
Naval Postgraduate School
Monterrey, CA 93940

Training Officer
Human Resource Management Center
NTC, San Diego, CA 92133

Navy Personnel R & D Center (5 copies)
Code 10
San Diego, CA 92152

Officer in Charge
Naval Submarine Medical Research Lab
Naval Submarine Base New London, Box 900
Groton, CT 06340

Officer in Charge (Code L5)
Naval Aerospace Medical Research Lab
Naval Aerospace Medical Center
Pensacola, FL 32512

Capt. Bruce G. Stone, U.S.N.
(Code N-33)
Director, Education & Training
Research & Program Development
Chief of Naval Education & Training
Staff
Naval Air Station, Pensacola, FL 32508

Dr. H. H. Wolff
Technical Director (Code N-2)
Naval Training Equipment Center
Orlando, FL 32813

Human Resource Management Center
Attachment
Naval Support Activity
c/o FPO New York, NY 09521
ATTN: TDC Nelson

Chief, Naval Technical Training
NAS Memphis (75)
Millington, TN 38128

Other

Division Director for Social Science
National Science Foundation
1800 G St., N. W.
Washington, DC 20550

Mr. Luigi Petrullo
2431 N. Edgewood St.
Arlington, VA 22207

Additions
to Distribution List

Cdr. Anthony C. Cajka, U. S. N.
Department of the Navy
Human Resource Management Center
Washington, DC 20370

Dr. C. Brooklyn Derr
Associate Professor, Code 55
Naval Postgraduate School
Monterey, CA 93940

Additions (continued)

Dr. Robert A. Zawecki
Assistant Professor of Behavioral
Sciences
6457B
United States Air Force Academy
USAFA, CO 80840

Captain E. L. Johnson, U.S.N.
Office of the Chief of Naval Operations
(OP-009F)
Navy Department
Washington, DC 20350

Bureau of Naval Personnel
Research & Evaluation Division
Code: Pers-65
Washington, DC 20370

Human Resource Management Center, London
FPO, NY 09510

Human Resource Management
Center, Washington
Washington, DC 20370

Human Resource Management Center, Norfolk
5621-23 Tidewater Dr.
Norfolk, VA 23511

Human Resource Management
Center, San Diego
Naval Training Center
San Diego, CA 92133

Human Resource Management Center,
Pearl Harbor
FPO San Francisco, CA 96610

Human Resource Management School
Naval Air Station, Memphis (96)
Millington, TN 38054

Capt. Bruce Stone, U.S.N.
Director
Program Development Div. (Code N-35)
Chief of Naval Education & Training
Naval Air Station
Pensacola, FL 32508

Mr. Keith Taylor
Office of Civilian Manpower Management
(Code 21)
Navy Department
Washington, DC 20390

Capt. Charles Baldwin, U. S. N.
Bureau of Naval Personnel
(Code 6a2)
Washington, DC 20370

Mr. Joel Ellermeier
Navy Personnel R & D Center
Code 308
San Diego, CA 92152

Office of Naval Research
(Code 200)
Arlington, VA 22217

ARI Field Unit-Leavenworth
P. O. Box 3122
Fort Leavenworth, KS 66027

Mr. Richard T. Mowday
College of Business Administration
University of Nebraska, Lincoln
Lincoln, NB 68588

Dr. William E. Gaymon
American Institutes for Research
3301 New Mexico Avenue, NW
Foxhall Square
Washington, DC 20016

Lt. Cdr. L. O. Milam
SMC 2165
Naval Postgraduate School
Monterey, CA 93940

Eugene F. Stone
Assistant Professor of Management
Dept. of Administrative Sciences
Purdue University
West Lafayette, IN 47907

Navy Material Command
Employee Development Office
Code SA-65
Room 150 Jefferson Plaza, Bldg. #2
1429 Jeff Davis Highway
Arlington, VA 20360

Headquarters, Forces Command
AFPE - HR
Ft. McPherson
Georgia 30330

Additions (continued)

Captain Joseph Weker
Department of the Army
Headquarters, 32D Army
Air Defense Command
APO N. Y. 09175

Edmund D. Thomas
(Code 307E7)
Navy Personnel Research
and Development Center
San Diego, CA 92152

Dr. Robert L. Holzbach, Jr.
Department of the Navy
Navy Personnel Research
and Development Center
San Diego, CA 92152

Johannes M. Pennings
Graduate School of Industrial Admin.
Carnegie-Mellon University
Schenley Park
Pittsburgh, PA 15213

Personnel Research and
Development Center
U.S. Civil Service Commission
Bureau of Policies & Standards
Washington, D. C. 20415

Department of the Air Force
Air Force Institute of Tech. (AU)
AFIT/SLGR (Lt. Col. Umstot)
Wright-Patterson Air Force
Base, Ohio 45433

Dr. John A. Drexler, Jr.
Battelle Human Affairs
Research Center
4000 N.E. 41st Street
Seattle, Washington 98105

Dr. Douglas T. Hall
Earl Dean Howard Professor
and Chairman
Dept. of Organizational Behavior
Graduate School of Management
Northwestern University
Evanston, Illinois 60201